ANALYSIS OF POLYMERISATION / AGGREGATION PROCESSES BY NIR CHEMICAL IMAGING AND FTIR SPECTROSCOPY

Idoia Martí Aluja

Dipòsit Legal: T. 1427-2013

Idoia Martí Aluja

ANALYSIS OF POLYMERISATION/AGGREGATION PROCESSES BY NIR CHEMICAL IMAGING AND FTIR SPECTROSCOPY

PhD. THESIS

Supervised by Prof. María Soledad Larrechi

Department of Analytical Chemistry and Organic Chemistry

UNIVERSITAT ROVIRA I VIRGILI

Tarragona, 2013

UNIVERSITAT ROVIRA I VIRGILI
DEPARTAMENT DE QUÍMICA ANALÍTICA I QUÍMICA ORGÀNICA

Campus Sescelades
Carrer Marcel·lí Domingo s/n
43007 Tarragona
Tel. +34 977 55 97 69
Fax. +34 977 55 84 46
e-mail: secqaqo@urv.cat

Dr. María Soledad Larrechi, Professor at the Department of Analytical Chemistry and Organic Chemistry at Universitat Rovira i Virgili

CERTIFIES:

That the PhD. Thesis entitled: **"Analysis of polymerisation / aggregation processes by NIR chemical imaging and FTIR spectroscopy"** submitted by Idoia Martí Aluja to achieve the degree of Doctor of Philosophy in Chemistry, has been carried out under my supervision at the Department of Analytical Chemistry and Organic Chemistry from this university, and fulfils the requirements to be eligible for the European Doctorate Award.

Tarragona, 27th September 2013

Prof. María Soledad Larrechi

Qualsevol nit pot sortir el sol

Jaume Sisa

I a vegades una tonteria de sobte ens indica que ens en sortim, Manel

De fet, la vida és això. Un cúmul de tonteries, petits moments i petits records que ens van articulant com a persones. I tots aquests petits moments van sempre acompanyats de persones petites però a la vegada grans, que donen un sentit a tot plegat. Al llarg d'aquests anys de la meva vida, he tingut la gran sort d'estar envoltada de gent fabulosa a qui m'agradaria donar gràcies pel simple fet de ser-hi.

Aquesta tesi no hagués estat possible sense l'ajuda, la guia i el suport constant de la meva directora de tesi, la Professora M. Soledad Larrechi. Marisol, no sabes cuánto he aprendido de ti. Profesionalmente me has ayudado a crecer como investigadora, a ser constante y a ver la investigación con unos ojos más humanos. Personalmente...he aprendido que lo importante no es la meta, sino el camino que recorremos. Gracias por darme la oportunidad de aprender y crecer a tu lado.

La resta de membres del grup de Quimiometria, Qualimetria i Nanosensors per la seva companyia i haver-me dedicat (com a mínim) un somriure pel passadís. M'he enriquit moltíssim amb tots vosaltres, he après diferents maneres de viure, de pensar i m'emporto una maleta plena d'experiències i molt bons records. Santi gràcies pel teu ajut i paciència, per estar sempre disposat a solucionar els meus problemes tècnics (sóc una experta en desastres informàtics).

Durant aquests anys he tingut la gran sort de poder fer una estada a la University of Copenaghen amb el Prof. Rasmus Bro. *Tak for at være hvem du er* and for teaching me that "Life is simple". Copenaghen is a magical city, full of people who welcomed me with open arms. Thanks, Tak, Grazie, Dank, Eskerrik asco, Obrigado, شـكرا, Gracias.. to make me feel like home. Skål! Colibrí, gracias por todo y por recordarme que todos los días sale el sol, chipirón.

A vegades esperem grans moments i deixem escapar els petits, i és què.. crec els grans moments són suma de petits instants que units configuren grans històries. Tinc grans històries compartides amb gent d'altres grups d'investigació. Començant pels meus companys orgànics: Miriam, Isma, Emma, Miriam S., Pep, Isidro, Foixinni, Quique, Isa, Maryluz, Camilo, Asta, Zeynep, Alev, Cristina L., Juanito... També els companys teòrics: Xipi, Alberto, Sonia,

Àlex, Núria A., Mireia, Joan Petit, Pablito, Pablo A., Jess, Gian Feliç, Magda, Marc, Gerard... Les zipi i zape inorgàniques Jessi i Eli, la Limón madrilenya (gracias por preocuparte por mi y dedicarme siempre unos minutos ☺), els companys futbolístics de Bioquímica. Des del primer fins a l'últim (i els que em deixo), tots heu compartit amb mi petits instants. Gràcies per fer-me sentir com una més i donar-me la oportunitat d'entrar en el vostre món. No sabeu com d'afortunada em sento.

Aquest camí no hagués estat igual de senzill sense els meus dos pajaritus preferits. M'és impossible resumir en quatre línies el que signifiqueu per mi i com d'especials sou ..

Miriam... Malgrat les experiències viscudes i compartides sempre tens aquesta capacitat de sorprendre'm i arrencar-me un somriure. Crec que no acabaria mai d'agrair-te tot el que has fet per mi i el suport incondicional que sempre m'has donat. Sempre tens la paraula i el consell adequat per cada situació, i això és una gran virtut que et fa especial. M'has ensenyat a saber valorar les persones i sobretot, a saber-me valorar a mi mateixa. Gràcies de tot cor.

Xavi... Tot i ser tant diferents hem sabut establir un equilibri, entendre'ns i ser molt bons amics. Moltes gràcies per saber treure'm el somriure quan ho he necessitat, saps com fer-me canviar el xip, i desconnectar de la vida fent un *click*. La vida t'ha obsequiat amb un do especial i màgic, una habilitat espectacular per reconfortar i fer sentir còmode a qualsevol. Sé que puc comptar amb tu pel que faci falta (no només pel tema de la boda, clar) i això no té preu per mi. Tak!

A la família vividora. Tant els que esteu per terres tarragonines, passant pels que esteu per Barcelona, Castelló, Londres, Berlin... sou molt especials. Una mica de manera casual, la vida em va col·locar en el vostre camí. I li estic molt agraïda perquè he conegut gent espectacular, molt vital i amb un cor enorme. *Chamaquitos*, gràcies per compartir amb mi tants somriures amb la vostra filosofia vividora i la vostra *buena onda*.

Aquest any ha estat un any de canvis, canvis que han anat acompanyats d'experiències positives que m'han ajudat a trobar-me a mi mateixa. De manera accidental m'he trobat vivint amb dues persones espectaculars en un barri que m'enamora. Frito tot i conviure tan poc temps.. parece como si nos

conociéramos de toda la vida. No perdis mai el bon humor i aquesta "xispa" que et fa únic i especial. *Punxing punxing*... Güeri he après tant amb i de tu.. Somiadora per naturalesa, has multiplicat per mil milions de vegades la meva felicitat. Gràcies per escoltar-me i estar allà quan ho he necessitat i per ajudar a alinear els meus "*shakres*" amb tots els consells i les conversacions sobre la vida a mitjanit (i sempre). Entrar al pis amb vosaltres ha estat un regal per mi. Quan sigui gran crearé un grup de fans vostres (Fans del *Piso de los Fritos*), del qual en seré la presidenta d'honor i amb molt orgull.

Lleida. Milers de moments compartits amb tots els amics de la terra ferma.. Moltes gràcies per dedicar algun minut de la vostra vida a fer-me feliç.

La meva vida no seria el mateix sense el suport de la meva família. Tiets and cousins.. gràcies pels bons moments i per ser com sou: genials. Padrins, la meva heroïna, la perseverança personificada. Tot i no entendre molt bé el què faig us dono les gràcies per estar allà i dedicar-me uns minuts per preguntar com em va "allò". Francesc i Emma gràcies per ser sempre allà i dedicar un instant de les vostres meravelloses vides per escoltar les meves preocupacions. Als meus pares... tot el que sóc i he après us ho dec a vosaltres. Aquesta tesi va dedicada a vosaltres, res no hagués estat possible sense vosaltres. Té sentit en vosaltres.

Gràcies també a tota la gent que em deixo. Hi han coses a la vida que no canvien mai, i una d'elles és la meva memòria de peix.

TABLE OF CONTENTS

3. Chemometric techniques 23

Chapter 1

General introduction

1. 1. Introduction

This doctoral thesis makes an *in situ* analysis of two processes that involve a physical state change from a liquid to a solid.

The first process studied is a polymerisation. In polymerisation processes, it is well-known that the physical and chemical properties of polymers are related not only to their composition and structure, but also the pathway of the polymerisation [1,2]. The most common way to study these processes is to analyse the sample at the beginning and at the end in different conditions [3]. To study the evolution of the polymer as it grows requires the process to be subject to continuous analysis, which cannot be done by the traditional techniques that study polymerisation because of the high cost of such commonly used techniques as nuclear magnetic resonance spectroscopy (NMR) [4-5]. This thesis reports the monitoring of a curing reaction between an enone-containing triglyceride and a diaminodiphenylmethane to obtain a polymer network. The enone-containing triglyceride reagent was obtained by an environmentally friendly procedure from high oleic sunflower oil use [6]. The use of reagents that come from vegetable oils presents an interesting alternative from the point of view of sustainability. Detailed information about the background to the reaction can be found in the introduction and the results sections of chapter 4.

The other process studied is protein aggregation, which involves a change in the structure of proteins that has important consequences for their bioactivity [7,8]. The kinetics of this process is affected by the medium conditions, and leads to the formation of different types of aggregate. This phenomenon is usually analysed by comparing the analytical information of the protein in its

native and aggregated form [7,8]. The traditional techniques used to study proteins and their processes are based on crystallographic studies, in which the sample must be a solid crystal [9-10]. This thesis focuses on the aggregation process of insulin, which is often used as a model. Chapter 5 presents detailed information of the process and its consequences in several fields.

The experimental developments that shape this doctoral thesis are the consequence of the working hypothesis that analytical methods based on infrared spectroscopy and chemometrics can give useful information for the *in situ* monitoring the physical and chemical changes of a process over time.

1. 2. Objectives

The overall objective of this thesis is to develop analytical methods based on infrared spectroscopy and chemometric techniques to analyse polymerisation and aggregation processes *in situ*. This overall objective has two separate sub-objectives:

1. To establish a methodology based on near-infrared chemical imaging (NIR-CI) and chemometric treatment to analyse a polymerization process. This objective requires:

    - Demonstrate whether NIR-CI is an appropriate technique for detecting the physical and chemical changes involved in a polymerisation process.

    - To use chemometric techniques suitable for image analysis.

2. To establish methodologies based on infrared spectroscopy and chemometric treatment for analysing the aggregation process of insulin in different medium conditions. To fulfil this specific objective we must:

   - Assess whether infrared spectroscopy is suitable for achieving this goal by making a detailed analysis of the spectral changes involved in the process.

   - Find a response that is useful for measuring the effect of the medium conditions on the process.

   - Plan the experimental part using strategies based on experimental design techniques to analyse the influence of biochemical variables.

   - Apply resolution curve methods to extract chemical information from the spectroscopic changes detected while the process is monitored.

   - Use other complementary techniques such as nuclear magnetic resonance (NMR), optical imaging or scanning electron microscopy (SEM) to obtain complementary information about how the process evolves.

1. 3. Structure of the thesis

This thesis is based on papers published in international journals. These papers have been edited so that the format is uniform. The contents have been structured in six chapters.

Chapter 1: Introduction. This chapter presents and defines the objectives and details of the various processes studied in the thesis.

Chapter 2: Instrumental techniques. This chapter describes certain theoretical aspects regarding the instrumental techniques used in this thesis. Essentially, it describes the basis of the two main techniques (near-infrared chemical imaging and infrared spectroscopy) and the complementary techniques used (nuclear magnetic resonance spectroscopy, optical imaging and scanning electron microscopy).

Chapter 3: Chemometric techniques. This chapter contains a brief description of the chemometric tools used to remove undesired contributions from the spectral data, to determine the chemical rank, to assess the homogeneity, to quantify the components of the aggregation process, to study the influence of biochemical variables on the aggregation and to find the response surface representative of the process.

Chapter 4: Analysis of a polymerisation process by NIR chemical imaging. This chapter describes the development of a method to monitor a polymerisation process *in situ* and to evaluate spatial distribution by near infrared chemical imaging (NIR-CI). It presents the results obtained as a scientific publication; therefore, it contains a brief introduction, the experimental procedure, the results and conclusions.

Chapter 5: Analysis of aggregation process by FTIR spectroscopy. This chapter focuses on establishing analytical methods to study the aggregation process of insulin through infrared spectroscopy and chemometric techniques. It is divided into two parts. The first part contains two scientific research papers on insulin aggregation in the presence of antiretroviral drugs. An overview of the process that was studied in the presence of these drugs is given in an introduction

section before the papers, which contain the results. The second part of the chapter studies simultaneously the values of some biochemical variables to determine how they influence insulin aggregation. We first studied the influence of these variables and the results are included in a research paper. Then, we modelled the aggregation time in function of the values of temperature, pH and ionic strength of a monomeric and a hexameric solution of insulin.

Chapter 6: Conclusions. This chapter contains the general conclusions of the thesis.

Appendix. This part of the thesis contains a list of abbreviations, a list of the published papers and the contributions made at meetings attended during this period.

1. 4. References

[1] Y. Cao, A. Andreatta, A. J. Heeger, P. Smith, *Polymer* **30** (1989) 2305 –2311.

[2] H. R. Kricheldorf, O. Nuyken, G. Swift, *Handbook of polymer synthesis*, Marcel Dekker, New York, 2005.

[3] I. M. Ward, *Structure and properties of oriented polymers*, Chapman & Hall, Cambridge, 1997.

[4] L. J. Mathias, *Solid state NMR of polymers*, Plenum Press, New York, 1989.

[5] Z-M. Huanga, Y-Z. Zhangb, M. Kotakic, S. Ramakrishna, *Compos. Sci. Technol*. **63** (2003) 2223 – 2253.

[6] L. Montero de Espinosa, J. C. Ronda, M. Galià, V. Cádiz, *J. Polym. Sci. A: Polym. Chem*. **46** (2008) 6843 – 6850.

[7] J. Brange, L. Andersen, E. D. Laursen, G. Meyb, E. Rasmussen, *J. Pharm. Sci.* **86** (1997) 517 –525.

[8] R. Malik, I. Roy, *Int. J. Pharm*. **413** (2011) 73 – 80.

[9] R. Nelson, M. R. Sawaya, M. Balbirnie, A. Ø. Madsen, C. Riekel, R. Grothe, D. Eisenberg, *Nature* **435** (2005) 773 – 778.

[10] M. Sunde, C. Blake, *Adv. Protein Chem*. **50** (1997) 123 – 124.

Chapter 2

Instrumental techniques

2. 1. Analytical techniques

Some aspects of the practical application of the instrumental techniques used in this thesis are described in this chapter. Two main techniques have been used throughout this thesis: near-infrared chemical imaging (NIR-CI) and Fourier transform infrared spectroscopy (FTIR). The fundamentals and the different types of acquisition modes of NIR-CI are briefly discussed in section 2.2. NIR-CI was used to control the homogeneity of a polymerisation process and the results are included in Chapter 4.

The basic background on infrared spectroscopy in diffuse reflectance mode is described in section 2.3. This technique was used to study the biochemical process of insulin aggregation. A short introduction is given to the absorption bands characteristic of proteins in that region (see section 2.4).

At the end of the chapter, brief mention is made of nuclear magnetic resonance (NMR), optical imaging and scanning electron microscopy (SEM), which were also used to obtain complementary information about how the processes evolved.

2. 2. Near-infrared chemical imaging (NIR-CI)

Near-infrared chemical imaging (NIR-CI) is an emerging technique that combines the advantages of digital imaging with spectroscopic measurements [1,2].

NIR chemical imaging instrumentation consists of a light source to illuminate the sample, a monochromator or filter system to disperse the light into wavelength bands, a detection system and a spectrometer camera to collect

the images [1, 3]. Spectral and spatial information is provided simultaneously. Three basic configurations are used, which differ in the way of acquiring the spatial information: point scan, line scan and plane scan [1,3, 4].

a) Point scan measurements

The camera device for this type of measurement is shown in Figure 2.1. A spectrum of a small region of the sample is measured from the diffuse reflectance or transmitted light at a single point on the sample. Then, the sample is displaced before a new spectrum is obtained. By moving the sample in two spatial dimensions (pixel positions) a complete hyperspectral image is acquired. The spatial size of the hyperspectral image is limited by the sample positioning hardware [1,2, 5,6].These instruments require long acquisition times, resulting in high resolution hyperspectral images. For this reason, they are widely used in stable samples. The main drawback of this type of instrument is that is difficult to reproduce the measurements, since the sample can only be positioned in the same place with great difficulty.

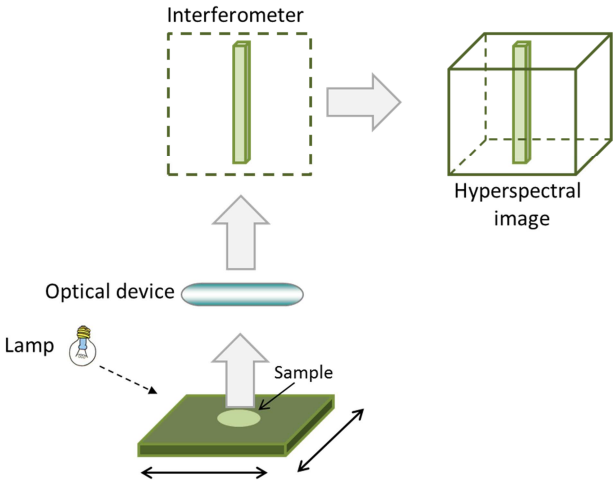


Figure 2. 1. Schematic representation of point scan instrumentation

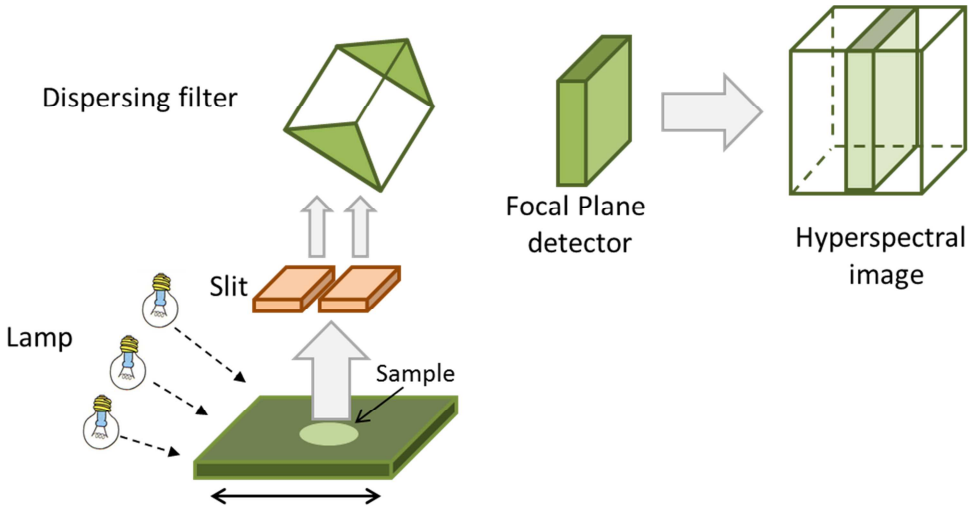


Figure 2. 2. Schematic representation of line scan instrumentation

b) Line scan measurements

A schematic representation of the second type of imaging device is shown in Figure 2.2. The main difference from the previous type is the detector, which is two dimensional and perpendicular to the sample. The sample is placed on a moving cell holder and imaged by a narrow line of radiation falling on the sample or by a narrow slit in the optical path leading to the detector [1,2,5]. Hyperspectral images are created by collecting sets of these matrices while moving the sample along the direction of the plane perpendicular to the detector line. The speed of image acquisition is less than in the previous case because it is limited only by the read speed of the camera. Commercial instruments can obtain images in minutes or even seconds, depending on the length of the spectral range selected.

c) Plane scan measurements

The third configuration is shown in Figure 2.3. In this configuration, the sample is illuminated by the radiation sources and after interaction with the sample; the reflected light is collected through an optical device (optics). Then, it is passed through a tuneable filter which sequentially changes each individual wavelength. An image at each individual wavelength is collected by measuring the intensity values in every pixel by the detector. The detector is a two-dimensional focal plane array (FPA), which is a series of cameras consisting of many thousands of individual detector elements (pixels).The hyperspectral image is formed by the combination of all individual images in the spectral range selected. The number of pixels in an image is determined by the FPA detector and the pixel size and scanned area is defined by the magnification optics [1,2,5].

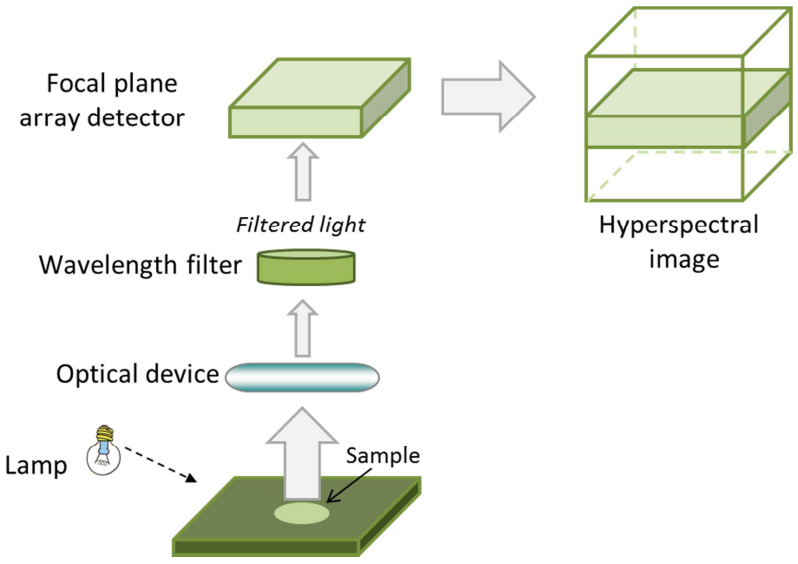


Figure 2. 3. Schematic representation of plane scan measurements

Plane scan devices have some advantages over the other types. First, all the components of these devices are fixed, which makes it more robust and applicable to *on-line* applications. Moreover, the acquisition time is shorter which make it possible to analyse less stable samples. The images captured throughout this thesis were acquired in this way, using a prototype located at the Universitat Autònoma de Barcelona (UAB).

In all cases, the data obtained is a three-dimensional data set, known as a hyperspectral data cube (**D**). An image is provided at individual wavelengths and a spectrum at each pixel (Figure 2.4). For this structure, the first two dimensions provide the spatial information (M, N) and the third one the spectral information (λ). Each unit of sample surface is called a pixel and contains one spectrum. Thus, the hyperspectral data cube (**D**) can be viewed as an array of spectra (one for each pixel) or a row of images at different wavelengths (the intensity values of every pixel are arranged next to each other) [1,7].

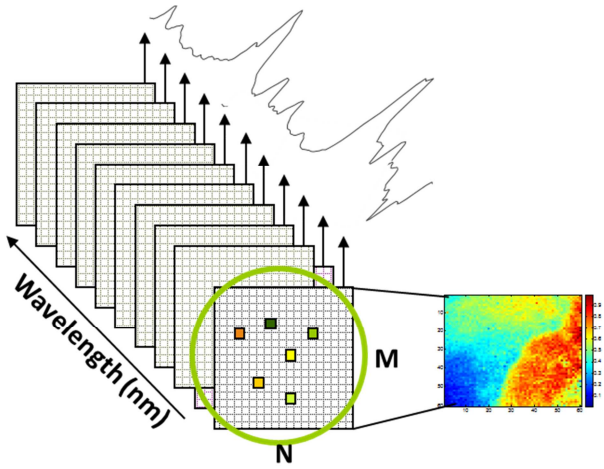


Figure 2. 4. Representation of the hyperspectral data cube

A hyperspectral cube consists of thousands of measurements from one sample. So it is evident that this data structure needs powerful statistical and mathematical treatment if the useful information is to be obtained. Chemometric tools enable all spectra in a data cube to be simultaneously analysed so that the desirable information can be extracted [2,5,6]. These tools will be explained in greater detail in Chapter 3.

2. 3. Fourier transform infrared spectroscopy-attenuated total reflection (FTIR-ATR)

Fourier transform infrared spectroscopy-attenuated total reflection (FTIR-ATR) is a valuable technique that has been widely used for many years to analyse chemical and biochemical processes [8,9]. It can be used to make qualitative and quantitative analyses of solid or liquid samples without any pre-treatment, which provides a spectrum quickly.

This study used attenuated total reflectance (ATR), which has the property of total internal reflection. In this type of measurement, a beam of infrared light is directed into the sample that is in contact with the ATR crystal and the light undergoes multiple internal reflections in the crystal. Some of the radiation is absorbed by the sample and the reflected radiation is returned to the detector. The resulting spectrum reflects the absorption of the functional groups of the sample that are active in the infrared region and constitutes a molecular fingerprint of the sample [8-10].

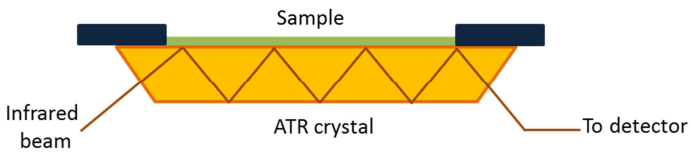


Figure 2. 5. Schematic representation of an ATR system

2.3.1. Absorption bands in proteins

The polypeptide units of proteins give rise to characteristic bands in the infrared region: namely, amide A, B and I-III [11,12]. Amide A (3100 - 3300 cm^{-1}) and amide B (3000 - 3100 cm^{-1}) bands appear due to the NH stretching vibration and do not depend on the backbone conformation [13]. However, they are sensitive to the hydrogen bond strength. Amide I and II are the two bands that provide most information for elucidating the secondary structure of proteins. The band that is most sensitive to the secondary structure is amide I (1600 – 1700 cm^{-1}, which arises from the C=O stretching vibrations of the peptide bonds. Thus, the position of this band is directly related to each type of secondary structure, which depends on the backbone structure [14,15]. The amide II band (1500-1600 cm^{-1}) complements the information obtained from the amide I band and is the result of the combination of the NH bending and the CN stretching vibration [16]. Amide III (1100-1330 cm^{-1}) is normally quite weak and occurs in a region of mixed vibrations but it can supplement the information obtained from the major bands [13,17]. The combination of the information in all of these regions provides insight into protein structure and evolution [11-18].

2. 4. ^{1}H Nuclear magnetic resonance spectroscopy (^{1}H NMR)

^{1}H Nuclear magnetic resonance (NMR) is a technique based on the magnetic properties of atomic nuclei. When a sample is exposed to a strong magnetic field, each ^{1}H absorbs electromagnetic radiation at a characteristic frequency as a function of the chemical environment of each ^{1}H [19, 20].

In this thesis, ^{1}H NMR has been used in the polymer field to elucidate the structure of the two forms in the tautomeric equilibrium of one of the reagents from vegetable oils, used as a starting product for a polymerisation process (Chapter 4). It was also used to confirm the disruption on the structure of insulin by antiretroviral drugs (Chapter 5).

2. 5. Microscopic imaging techniques

Microscopic imaging techniques are used in a range of fields not only to enhance and extract powerful information but also to complement information from other techniques. During this thesis, optical microscopy and scanning electronic microscopy (SEM) were used to obtain complementary information about the processes studied. A brief explanation of the background of these techniques is described below.

2. 5. 1. Optical microscopy

Optical microscopy is a well-established technique that provides a magnified view of a sample (Figure 2.6. a)). The sample is placed on a microscope slide made of glass and the visible light is passed through it. Then, this light passes through one or more objective lenses that produce enlarged images of a

sample. The real power or magnification of an image is proportional to the power of the objective lens used. The typical magnification is up to 1500x with a theoretical resolution limit of around 0.2 micrometres (200nm) [21, 22].

2. 5. 2. Scanning electron microscopy (SEM)

A scanning electron microscope (SEM) is a type of electron microscope that uses an electron beam to produce magnified images of a sample. The electrons interact with the sample and produce various signals that contain information about the sample surface and composition (see Figure 2.6. b)) [23].

SEM samples (or at least their surface) must be electrically conductive if they are to acquire a SEM image. In most cases, the sample does not conduct electricity, so they are coated with a thin layer of an electrically conducting material (such as gold, palladium, platinum, etc) [24].

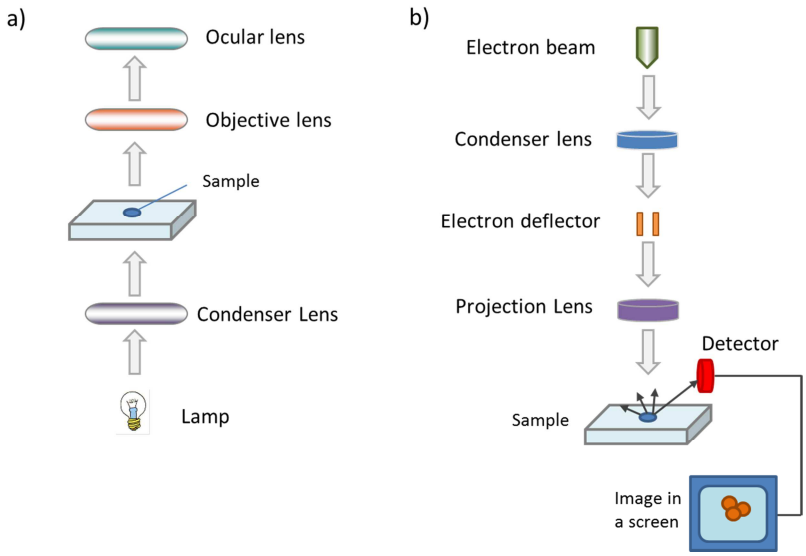


Figure 2. 6. Schematic representation of a) an optical microscopy; and b) a scanning electron microscopy (SEM).

Both techniques were used to characterise insulin aggregates formed at the end of the process (Chapter 5).

2. 6. References

[1] H. F. Grahn, P. Geladi, *Techniques and applications of hyperspectral image analysis,* WILEY-VCH, Chichester, 2007.

[2] J. M. Amigo, *Anal. Bioanal. Chem*. **398** (2010) 93-109.

[3] P. Geladi, K. Esbensen, S. Wold, *Anal. Chim. Acta* **191** (1986) 473–480.

[4] J. M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco, *Trends Anal. Chem*. **27** (2008) 696-713.

[5] R. Salzer, H. W. Siesler, *Infrared and Raman spectroscopic imaging*, WILEY-VCH, Weinheim, 2009.

[6] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, *Trends Anal. Chem.* **23** (2004) 70-79.

[7] J. M. Prats-Montalban, A. de Juan, A. Ferrer, *Chemometr. Intell. Lab.* **107** (2011) 1–23.

[8] J. C. Chalmers, P. R. Griffiths, *Handbook of vibrational spectroscopy*, John Wiley & Sons, Chichester, 2002.

[9] D. N. Sathyanarayana, *Vibrational spectroscopy: theory and applications*, New age international Ltd., New Delhi, 2004

[10] F. M. Mirabella, *Internal reflection spectroscopy: theory and applications*, Marcel Dekker Inc., NewYork, 1993.

[11] B. R. Singh, *Infrared analysis of peptides and proteins: principles and applications*, ACS, Washington DC, 2000.

[12] A. Barth, P. I. Hari, *Biological and biomedical infrared spectroscopy*, IOS Press BV, Amsterdam, 2009.

[13] A. Barth, C. Zscherpmide, *Q. Rev. Biophys*. **35** (2002) 369–430.

[14] P. I. Haris, F. Severcan, *J. Mol. Catal. B-Enzym*. **7** (1999) 207-221.

[15] J. Kong, S. Yu, *Acta Bioch. Bioph*. Sin. **39** (2007) 549-559.

[16] M. Carbonaro, A. Nucara, *Amino acids* **38** (2010) 679-690.

[17] P. C. Groβ, M. Zeppezauer, *J. Pharmaceut. Biomed*. **53** (2010) 29-36

[18] A. Dong , P. Huang , W. S. Caughey, *Biochem*. **29** (1990) 3303–3308.

[19] T. N. Mitchell, B. Costisella, *NMR - from spectra to structures: an experimental approach,* Springer-Verlag Berlin Heidelberg, Heidelberg, 2007.

[20] N. E. Jacobsen, *NMR spectroscopy explained: simplified theory, applications and examples for organic chemistry and structural biology,* John Wiley & sons, Hoboken, 2007.

[21] J. Mertz, *Introduction to optical microscopy*, Roberts & Company Publishers, Colorado, 2010

[22] B. Herman, J. J. Lemaster, *Optical microscopy: emerging methods and applications*, Academic Press Inc., San Diego, 1993.

[23] D. B. Williams, C. B. Carter*, Transmission electron microscopy*, Plenum Press, New York, 1996.

[24] P. Echlin, *Handbook of sample preparation for scanning electron microscopy and X-ray microanalysis*, Springer Science + Business Media, New York, 2009.

Chapter 3

Chemometric techniques

3. 1. Introduction

When bidimensional or tridimensional data are obtained from the analysis of a sample, such as the data obtained throughout this thesis, chemometric tools are required to extract useful information from them.

Chemometrics was defined in 1975 as the set of methods used to extract chemical information from a group of data. Massart *et al.* gave a complete and precise definition of this term: "Chemometrics is the chemical discipline that uses mathematical, statistical and other methods employing formal logic (a) to design or select optimal measurement procedures and experiments, and (b) to provide maximum relevant chemical information by analysing chemical data" [1].

In the following thesis, chemometric techniques have been used with different purposes:

- To remove undesired contributions from the spectral data.
- To determine the number of chemical compounds involved in the processes studied.
- To assess the homogeneity of the sample analysed.
- To quantify the components involved in the aggregation process of insulin.
- To study the influence of some biochemical variables on the aggregation process of insulin, and to find the response surface

representative of the relationship between the aggregation time and the values of the biochemical variables.

The basis of the chemometric techniques used for each purpose is included in the following sections.

3. 2. Unfolding of the hyperspectral data cube

The hyperspectral data cube acquired by chemical imaging techniques needs to be adapted to a two-dimensional array before any mathematical treatment is applied. This process is called unfolding and the hyperspectral data cube (**D**) is transformed into a two-dimensional matrix (**D**), called an unfolded data matrix. Each row from this last structure is the spectrum of one of the pixels and the total number of columns is the total number of pixels (Figure 3.1.) [2,3]. As the most important feature of hyperspectral data cube is the spatiality, once the data cube is processed the data matrix can be refolded into a three-dimensional structure to retain the pixel location.

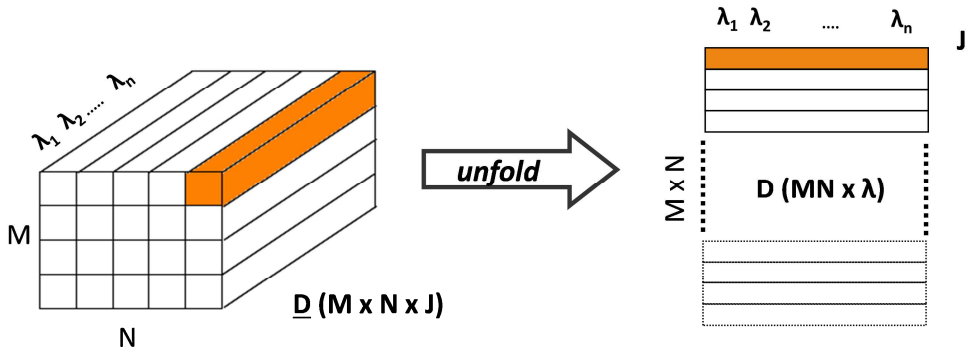


Figure 3. 1. Schematic representation of the hyperspectral data cube unfolding

3. 3. Chemometric tools for removing undesirable contributions

Spectroscopic data can contain undesirable chemical and non-chemical contributions. The non-chemical information may be originated from the sample (e. g. uneven surface) and /or by instrumental variations or noise. All these effects can hamper the analysis. Therefore, it is crucial to select the appropriate pre-processing algorithm to suppress the undesired information and enhance the chemical information [2,3,4].

The pre-treatment techniques employed throughout this thesis are:

- *Savitzky-Golay smoothing average*: it is used to reduce the spectral noise and improve the signal/noise ratio. This algorithm uses a polynomial window and determines the mean of its consecutive values. Then, the window is displaced one point through the spectra and the mean is calculated. This procedure is repeated through the spectra and midpoints are placed one behind the other, obtaining a new smoothed spectrum [5].

- *Standard normal variate (SNV):* it is used to correct baseline drifts caused by light scattering. This method transforms each spectrum into a spectrum with mean intensity 0 and standard deviation 1, following Eq. (2.1):

$$Abs_j^{SNV} = \frac{Abs_j - \overline{Abs}}{S} \qquad (2.1)$$

where Abs_j is the original absorbance at *j* wavelength, $\overline{Abs}$ is the mean absorbance of the spectrum and *S* is the standard deviation [6].

3. 4. Chemometric tools for assessing the rank by factor analysis

3. 4. 1. Principal component analysis (PCA)

Principal component analysis (PCA) is one of the most used exploratory techniques in a variety of fields. It is a variable reduction technique that seeks to extract the most information from a data matrix and reduce it to a few variables, called principal components (PCs). These PCs are new uncorrelated linear combinations of the original variables and they are orthogonal. The first PC describes the largest variation in the data matrix, the second PC the second largest and so forth [1].

PCA decomposes the experimental data matrix **D** (*I* x *J*) into the product of two matrices, according to:

$$\mathbf{D} = \mathbf{T}\mathbf{P}^{T} + \mathbf{E} \qquad (2.2)$$

where **T** (*I* x *F*) and $\mathbf{P}^{T}$ (*J* x *F*) account for the scores and loadings matrices for F principal components (PCs), respectively. **E** (*I* x *J*) is the residual matrix, which contains the information not explained by $\mathbf{T}\mathbf{P}^{T}$. The scores matrix (**T**) contains information about the samples, and $\mathbf{P}^{T}$ is the loadings matrix. The rows of this matrix are called principal components and they are linear combinations of the *J* original variables so they provide information about the variables that better describe the variation on the original data matrix (**D**).

PCA can be performed using various algorithms. Among them, singular value decomposition (SVD) has been used in this thesis to analyse the number of variability sources present in the experimental data [1]. This algorithm shows that a matrix **D** of dimensions (*I* x *J*) can be decomposed into the product of

three matrices **U** (*I* x *K*), **W** (*K* x *K*) and $\mathbf{P}^T$ (*K* x *J*), where *K* is the lowest value between *I* and *J*.

$$\mathbf{D} = \mathbf{UWP}^T \qquad (2.3)$$

The diagonal elements of matrix **W** are positive values, whereas the elements outside the diagonal are zero. The product between matrices **U** and **W** gives the scores matrix mentioned above, whereas $\mathbf{P}^T$ is the loadings matrix [1].

The diagonal elements of matrix **W** are called *eigenvalues* or *singular values*, and they are sorted in increasing order. Information about the number of factors or components that have relevant information about the system studied can be obtained by representing the *singular values* as a function of the number of factors. Factors with low *singular values* usually contain information related to instrumental noise. The combination of this information with the loadings leads to obtain the chemical rank of a data matrix, which is assumed to be the number of significant compounds.

3. 4. 2. Fixed size image window-evolving factor analysis (FSIW-EFA)

Fixed size imaging window-evolving factor analysis (FSIW-EFA) is an adaptation for images of the well-known fixed size moving window-evolving factor analysis (FSMW-EFA) [2,3,7]. FSIW-EFA mimics its procedure to calculate the rank of a matrix. The general steps are shown in Figure 3.2. and they are commented below.

1. Selection of the window size that will enclose a central pixel and its neighbouring pixels (pixel area).

2. This sub-set of the hyperspectral cube (pixel area) is unfolded and a data matrix is obtained. Then, the singular value decomposition (SVD) is performed to evaluate its rank.

3. The window is moved one pixel at a time, and SVD is carried out for each pixel area, until the entire image is covered.

4. The *singular values* (SV) of each unfolded matrix are displayed together in singular value plots. This step is essential to select the threshold value, which will mark the limit between significant information and noise. The number of components equals the number of *significant values* present for each pixel area.

5. Each pixel area contains the individual rank, which is refolded to recover the initial dimensions of the image. Then, the local rank map can be represented, which shows the number of components that are present in the image.

To sum up, FSIW-EFA is a local rank analysis algorithm that uses the spatial structure of the image (three-dimensions) and also preserves the surrounding of each pixel. The local rank maps are easy to interpret and offer an excellent synopsis of the image complexity [7,8].

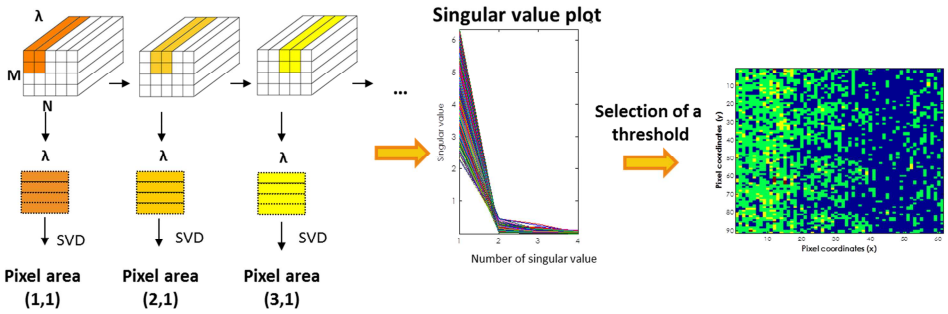


Figure 3. 2. Visual representation of the procedure followed for FSIW-EFA algorithm

3. 5. Chemometric tools for assessing the homogeneity of a sample using hyperspectral images

An easy way to obtain qualitative visual information about the distribution of pure species in an image is to use correlation maps. First of all, the images must be unfolded as is represented in Figure 3.1. to get the unfolded matrix. The procedure for obtaining correlation maps is to calculate the correlation coefficient between the spectrum representative of the pure species and each spectrum of each single pixel in the image (rows in the unfolded matrix), as follows:

$$corrcoef = \frac{\sum_\lambda x_{mnj} y_j}{\sqrt{\sum_\lambda x_{mnj}^2 \sum_j y_j^2}} \qquad (2.4)$$

where x_{mnj} is the spectrum measured in the *ij*th pixel for each wavelength and y_j is the spectrum representative of the pure species. Correlation coefficients are obtained for each pixel, and they are re-arranged to form a correlation map or correlation surface (see Figure 3.3). This map provides an overview of the

similarity between the spectra, with 1 being the value for the maximum similarity [9,10].

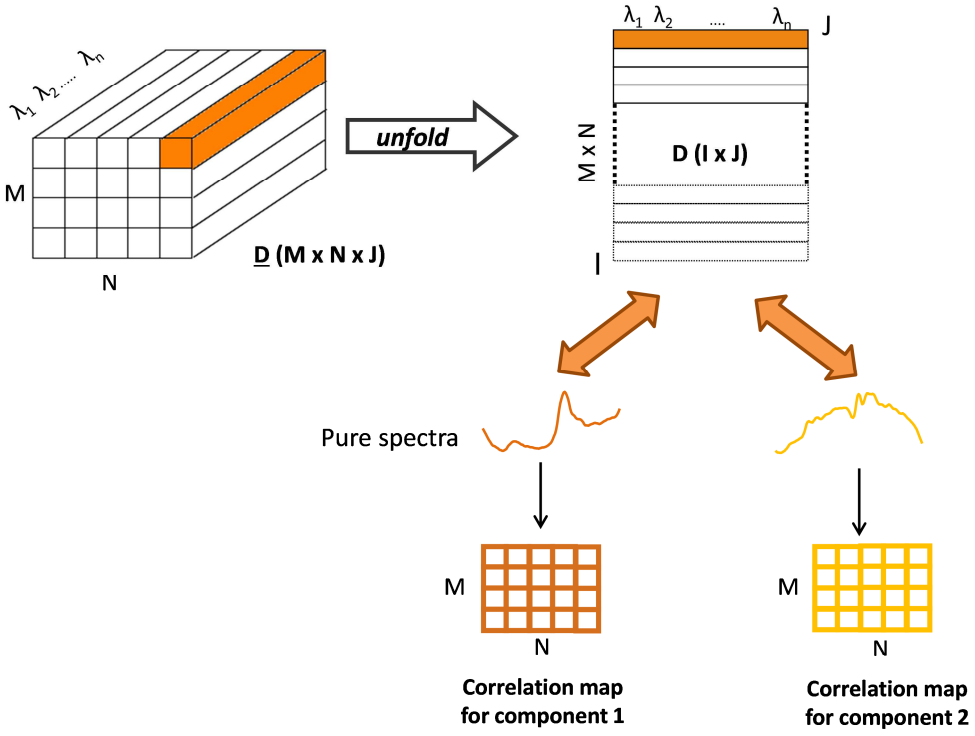


Figure 3. 3. Schematic representation of the steps involved in the correlation maps

3. 6. Chemometric tools for quantification: multivariate curve resolution-alternating least squares (MCR-ALS)

The main goal of multivariate curve resolution-alternating least squares (MCR-ALS) is to isolate, solve and quantify the variability sources present in a data matrix. This method is well-known and widely used to solve problems in a range of evolving systems [11,12].

Assuming that the experimental data follow a bilinear model, MCR-ALS decomposes the data matrix (**D**) into the product of two matrices according to:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E} \qquad (2.5)$$

where **D** (*I* x *J*) is the original data matrix, **C** (*I* x *F*) contains the concentration profiles of the different species and $\mathbf{S}^{\mathbf{T}}$ (*F* x *J*) is their related pure spectra. **E** (*I* x *J*) is the experimental error unexplained by the resolution model. *I* is the number of analysed spectra, *J* is the number of variables recorded and *F* is the number of species present in the system studied [11-13].

MCR-ALS iteratively solves Eq. (2.5), following these stages (see Figure 3.4.):

*a) Determination of the number of compounds in **D***

First, the number of components that significantly contribute to the response variable needs to be determined. There are several techniques for doing this. Principal component analysis (PCA) [14,15], evolving factor analysis (EFA) [1,16,17] and singular value decomposition (SVD) [1] are the most widely used. In this thesis, SVD was used and the number of retained factors was estimated by examining the magnitude of the *singular values*. Factors with *singular values* larger or equal to the experimental error were rejected.

*b) Construction of an initial estimate of **C** or $\mathbf{S}^{\mathbf{T}}$*

Once the number of species involved in the system is known, initial estimations of the concentration (**C**) or the spectral profiles ($\mathbf{S}^{\mathbf{T}}$) are constructed. Techniques such as evolving factor analysis (EFA) [16,18,19] or the simple-to-use interactive self-modeling mixture (SIMPLISMA) [20,21] provide initial estimates of **C** or $\mathbf{S}^{\mathbf{T}}$. In this thesis, EFA profiles were used as initial estimates as they are the most suitable choice for time-series data.

c) Optimisation process by alternating least squares (ALS)

The initial estimates for **C** and $\mathbf{S}^T$ are optimised by iteratively solving Eq. (2.5) using alternating least squares (ALS) [22,23] and these steps:

1. The spectral profiles are obtained from the estimation of concentration profiles, according to:

$$\mathbf{C}^+ \mathbf{D}^* = \mathbf{C}^+ \mathbf{C}\mathbf{S}^T = \mathbf{S}^T \qquad (2.6)$$

   where $\mathbf{C}^+$ is the pseudoinverse of matrix **C** and $\mathbf{D}^*$ is the experimental matrix gained from PCA for the selected number of PCs. $\mathbf{S}^T$ is the pure spectra matrix, calculated by least squares.

2. Then, a new estimate of the concentration profiles is obtained from the spectral profiles obtained in the previous step:

$$\mathbf{D}^* (\mathbf{S}^T)^+ = \mathbf{C} (\mathbf{S}^T) (\mathbf{S}^T)^+ = \mathbf{C} \qquad (2.7)$$

Both steps are repeated until the difference between the residual of one iteration and the next is less than a fixed value (typically of the order of 0.1).

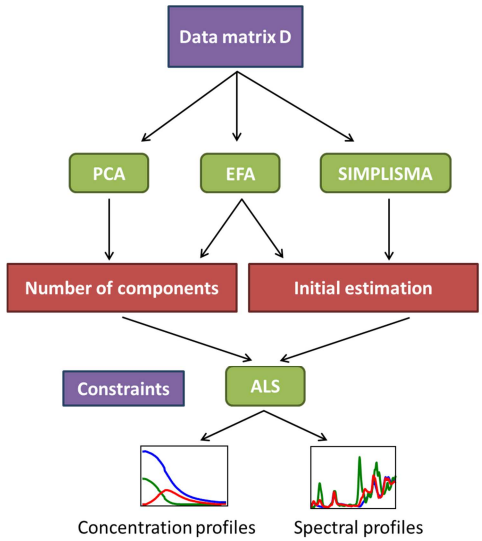


Figure 3. 4. Scheme of the resolution process followed by MCR-ALS

This iterative process can be improved by using additional information based on previous knowledge of the systems [24, 25,26]. This information is included in the algorithm by the implementation of constraints. The most commonly used are:

- *Non-negativity*, the concentration and/or spectral profiles are forced to have positive or zero values.

- *Unimodality*, a single peak (one maximum) is imposed in each response profile.

- *Closure*, the sum of the concentrations of all species involved in the system is forced to be constant at each stage of the reaction.

- *Selectivity*, used when zones are known to be totally selective for one of the species. It is the same as restricting the concentrations or signals of the other components to zero.

There are several parameters for evaluating the goodness of the adjustment made by MCR-ALS. Specifically, the lack of fit (LOF) of the model and explained variance (R^2) are the parameters that are most used for this purpose:

$$LOF\ (\%) = \ 100 \cdot \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \qquad (2.8)$$

$$R^2(\%) = \left(1 - \frac{\sum_{i,j} d_{ij}^2 - \sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}\right) \cdot 100 \qquad (2.9)$$

where d_{ij} is an element of the input data matrix **D** and e_{ij} is the related residual obtained from the difference between the experimental data matrix (**D**) and the MCR-ALS reproduction. The closer the values of LOF and R^2 to zero and one, respectively, the better the fit of the experimental data [24-27].

On the other hand, if information is available about the real profiles, the quality of the recovered profiles can be assessed by the similarity coefficient (r). This parameter represents the correlation between the recovered profiles and the real profiles:

$$r = cos\ \gamma = \frac{s_i^T \hat{s}_i}{\|s_i\| \cdot \|\hat{s}_i\|} \qquad (2.10)$$

where γ is the angle of the vectors associated with the profile recovered by MCR-ALS ($\hat{s}_i$) and the real profile (s_i), for each species *i*.

3.6.1. MCR-BANDS

As mentioned above, the main goal of MCR-ALS is to obtain a meaningful chemical solution, which describes the real profiles of the different components present in the system [28]. However, there are an infinite number of solutions as there are infinite linear combinations of **C** and $\mathbf{S}^T$, which produce the same

result (**D**) when multiplied. This rotational ambiguity is defined by the following equation:

$$\mathbf{D}^{*} = \mathbf{C}_{old}\,\mathbf{S}_{old}^{T} = (\mathbf{C}_{old}\,\mathbf{T})\,(\mathbf{T}^{-1}\,\mathbf{S}_{old}^{T}) = \mathbf{C}_{new}\,\mathbf{S}_{new}^{T} \qquad (2.11)$$

According to Eq. (2.11), any invertible matrix **T** gives a new set of equivalent solutions. The number of possible solutions can be reduced using constraints arising from prior knowledge of the system.

One way to evaluate the rotational ambiguity in MCR-ALS is to calculate the feasible bands in all possible solutions. For a particular species profile, the set of feasible solutions under constraints defines a range or band of feasible solutions, and this band may be delimited by maximum and minimum band boundaries. In this thesis, the boundaries were defined using the MCR-BANDS method, which obtains them from a constrained non-linear optimisation of an objective function f_n (**T**) that depends on the rotation matrix **T** (Eq. (2.11)) [28,29,30]. This function is scaled between 0 and 1 and estimates the contribution of each component to the overall signal profile.

$$f_n(\mathbf{T}) = \frac{\|\mathbf{c_n s_n^T}\|}{\|\mathbf{CS^T}\|} \qquad (2.12)$$

3. 7. Experimental design techniques

Experimental design techniques provide tools and strategies for:

a) Analysing the influence of the factors (variables) and their interaction with the experimental response.

b) Modelling the response surface that relates the experimental response and the values of the factors.

The techniques used to achieve these goals are briefly described below.

3.7.1. Full factorial design

Full factorial design techniques enable the influence of the factors and their interactions on the experimental response. The experimental set-up consists of all the possible combinations between the variables and their levels. The experimental matrix contains the codified values of variables (or factors) for each experiment. In general, a design with *k* factors that are considered at *m* levels contains m^k experiments. In this thesis, full factorial design was used at two levels [31,32].

The influence of the factors and their interactions on the response is assessed by the significance of coefficient b_i, calculated from the model matrix linked to the 2^k design:

$$b_i = \frac{\sum y(+)}{n} - \frac{\sum y(-)}{n} \qquad (2.13)$$

where $\sum y(+)$ and $\sum y(-)$ are the sums of the responses where factor *i* is at its high (+1) and low (-1) level in the model matrix related to the experimental matrix, and *n* is the number of experiments. This coefficient can be calculated for each factor or for each interaction (b_{ij}), included in the model matrix.

To assess the significance of the coefficients b_i and b_{ij}, their variability was compared with an estimate of the experimental error by Analysis of Variance (ANOVA). The effect of the studied factor is important when both values differ statistically [1,33].

3.7.2. Modelling of a response: response surface methodology

The mathematical model that relates the experimental response as a function of the values of the factors was found using response surface methodology [34]. The basis of this strategy is briefly described below.

The order of the function that determines the relationship between the response and the factors is sometimes unknown, and it can be modelled by a polynomial function. The first step consists of establishing a first-order model to find an approximation of the true relationship between the response and the variables, following Eq. (2.14):

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i<j} \sum b_{ij} x_i x_j + \varepsilon \qquad (2.14)$$

where y is the response value, b_0, b_i and b_{ij} are the coefficients of the model and x_i and x_j are the codified values of the variables. b_0 is a constant term which estimates the response when the values of all factors are in the centre points of the model, b_i are the linear coefficients which describe the sensitivity of the response to the variations in the corresponding factors, and b_{ij} are the crossproduct coefficients which enable to establish the first-order response surface.

If the model is not valid (see section 3.4.6.), more experiments are needed to establish a second-order model. Although there are several ways of achieving this goal, central composite design was used in this thesis. Its main advantage is that it makes possible to use the experiments from the first-order model. This strategy includes embedded factorial designs, central points and axial points for estimating the curvature of the model, as follows:

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i=1}^{k} b_{ii} x_i^2 + \sum_{i<j} \sum b_{ij} x_i x_j + \varepsilon \qquad (2.15)$$

where y, b_0, b_i, b_{ij}, x_i and x_j have the same meaning as in the previous case and b_{ii} is a new coefficient that corresponds to the quadratic terms that describe the curvature.

3.7.3. Validation of the model

An experimental way to evaluate whether a model is representative of the real information is to perform replicates on the centre point. If the difference between the average of the responses at the central point ($\overline{y_C}$) and the average of the responses of the experiments in the experimental plan ($\overline{y_F}$) is small, the centre points lie on or near the plane passing through the factorial points and there is no quadratic curvature. On the other hand, if this difference is large, then quadratic curvature is present.

To decide when this difference is high and verify the presence of curvature, the sum of squares for pure quadratic curvature ($SS_{pure\ quadratic}$) can be calculated as follows:

$$SS_{pure\ quadratic} = \frac{n_F n_C (\overline{y_F} - \overline{y_C})^2}{n_F + n_C} \qquad (2.16)$$

where n_F is the number of points in the factorial design and n_C is the number of replicates at the central point.

This value ($SS_{pure\ quadratic}$) is compared with the $SS_{residual}$ obtained with the replicates of the central point by means of an F-test. There is no quadratic curvature when both values are similar. On the other hand, if they are not, the model is described by a quadratic curvature so that the model has to be represented by a second-order function [31].

The prediction error can be evaluated with external samples. Other experiments within the experimental domain must be performed.

Then, the presence of bias can be assessed by the join statistical test of the regression between the measured and predicted responses. If there are no significance differences between the two values (measured and predicted), the values of the slope and intercept of the straight regression line should fall within the elliptical confidence region (normally α=0.05) centred on the theoretical values of 1 slope and 0 intercept.

3.8. References

[1] D. L. Massart, B. Vandeginste, L. Buydens, S. de Jong, P. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A,* Elsevier, Amsterdam, 1997.

[2] H. Grahn, P. Geladi, *Techniques and applications of hyperspectral image analysis,* WILEY-VCH, Chichester, 2007.

[3] R . Salzer, H. W. Siesler, *Infrared and Raman spectroscopic imaging*, WILEY-VCH, Weinheim, 2009.

[4] J. M. Prats-Montalban, A. de Juan, A. Ferrer, *Chemom. Intell. Lab.* **107** (2011) 1 –23.

[5] A. Savitzky, M. E. Golay, *Anal. Chem.* **36** (1964) 1627 –1639.

[6] R. J. Barnes, M. S. Dhanoa, S. J. Lister, *Appl. Spectrosc.* **113** (1989) 1849 – 1854.

[7] A. de Juan, M. Maeder, T. Hancewiczc, R. Tauler, *Chemom. Intell. Lab. Syst.* **77** (2005) 64 – 74.

[8] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, *Trends Anal. Chem*. **23** (2004) 70 – 79.

[9] J. M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco, *Trends Anal. Chem.* **27** (2008) 696 –713.

[10] J. M. Amigo, *Anal. Bioanal. Chem.* **398** (2010) 93 – 109.

[11] R. Tauler, A. Smilde, B. R. Kowalski, *J. Chemometr.* **9** (1995) 31 – 58.

[12] R. Tauler, B. R. Kowalski, S. Flemming, *Anal. Chem.* **65** (1993) 2040 – 2047.

[13] T. Azzouz, R. Tauler, *Talanta* **74** (2008) 1201 – 1210.

[14] E. R. Malinowski, *J. Chemometr.* **13** (1999) 69 – 81.

[15] D. L. Massart, B. Vandeginste, S. N. Deming, Y. Michotte, L. Kaufman, *Chemometrics: a textbook. Part B*, Elsevier, Amsterdam, 1988.

[16] M. Maeder, *Anal. Chem.* **59** (1987) 527 – 530.

[17] H. Gampp, M. Maeder, C. J. Meyer, A. D. Zuberbühler, *Talanta* **33** (1986) 943 –951.

[18] E. R. Malinowski, *J. Chemometr.* **10** (1996) 273 – 279.

[19] A. C. Whitson, M. Maeder, *J. Chemometr.* **15** (2001) 475 – 484.

[20] W. Widing, N. B. Gallagher, J.M. Shaver, B. M. Wise, *Chemom. Intell. Lab. Syst.* **77** (2005) 85 – 96.

[21] W. Widing, *Chemometr. Intell. Lab.* **36** (1997) 3 – 16.

[22] R. Tauler, A. Izquierdo-Ridorsa, E. Cassassas, *Chemom. Intell. Lab. Syst*. **38** (1997) 293 – 300.

[23] A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou, R. Tauler, *Chemom. Intell. Lab. Syst*. **18** (1997) 183 – 196.

[24] R. Bro, S. De Jong, *J. Chemometr.* **11** (1997) 393 – 401.

[25] R. Tauler, E. Casassas, *Anal. Chim. Acta* **248** (1991) 447 –458.

[26] R. Gargallo, F. Cuesta-Sánchez, D. L. Massart, R. Tauler, *Trends Anal. Chem.* **15** (1996) 279 – 286.

[27] A. de Juan, Y. Vander-Hieden, R. Tauler, D. L. Massart, *Anal. Chim. Acta* **346** (1997) 307 – 318.

[28] C. Ruckebusch, L. Blanchet, *Anal. Chim. Acta* **765** (2013) 28 – 36.

[29] J. Jaumot, R. Tauler, *Chemom. Intell. Lab. Syst.* **103** (2010) 96 – 107.

[30] X. Zhang, R. Tauler, *Anal. Chim. Acta* **762** (2013) 25 – 38.

[31] D. C. Montgomery, *Design and analysis of experiments,* Wiley, Hoboken, 2001.

[32] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, A. Nyström, J. Pettersen, R. Bergman, *Chemom. Intell. Lab. Syst.* **42** (1998) 3 – 40.

[33] R. Mead, *The design of experiments: statistical principles for practical applications*, Cambridge University Press, Melbourne, 1988.

[34] G. Hanrahan, K. Lu, *Crit. Rev. Anal. Chem.* **36** (2006) 141 – 151.

Chapter 4

Analysis of a polymerisation process by NIR chemical imaging

4.1. Introduction

Chapter 4 aims to describe the development of a fast method based on near infrared chemical imaging (NIR-CI) to monitor a polymerisation process *in situ* and to evaluate their spatial distribution.

The physical and chemical properties of polymers are related not only to their composition and structure, but also to the spatial distribution of the polymer chain during the polymerisation process [1,2]. Although NMR imaging is a technique often used to obtain spatial information about the polymer, it is not easily available and the results are complex [3,4]. Therefore, the challenge is to assess the possibility of using other simpler techniques to obtain similar information.

In the present thesis, NIR-CI was chosen because of the low molar absorptivity of the species in this region [5-8]. This enables the polymerisation processes to be analysed *in situ*, providing spectral changes that happen over time and spatial information of how they evolve.

Developing polymer materials from renewable resources is becoming increasingly interesting from the point of view of sustainability [9]. Vegetable oils are regarded as the most important class of renewable raw materials, and are mainly used for synthesising of plant-oil based polymers [10,11]. They are mainly composed of triglycerides, which normally crosslink with other chemical agents. While the crosslinked polymers are forming, the system evolves from a liquid state to an insoluble solid state [12].

Crosslinked polymers are usually synthetized using epoxidized vegetable oils with conventional aromatic diamines as curing agents [13,14]. An interesting

alternative to these epoxidized vegetable oils is to use an enone-containing triglyceride as starting the oil, which can undergo an aza-Michael reaction to yield the polymer [15,16].

This chapter focuses on the polymerisation reaction between an enone-containing triglyceride (ENO) and diaminodiphenylmethane (DDM) [15]. Figure 4. 1. shows a schematic representation of the experimental procedure performed. Both products were solids at room temperature so they were weighed and mixed in a ceramic pestle until a homogeneous mixture was obtained. The crosslinking process was catalysed by borontrifluoride monoethanolamine complex (BF_3 · MEA), which was added to the mixture before heating. The mixture was placed in a homemade circular crystal cell and then in an oven near the instrument at T=95 ºC. The total reaction time and other information about the experimental procedure and instrumental measurements can be found in the results section in this chapter. This section also includes a more detailed representation of the reactions involved in the polymerisation.

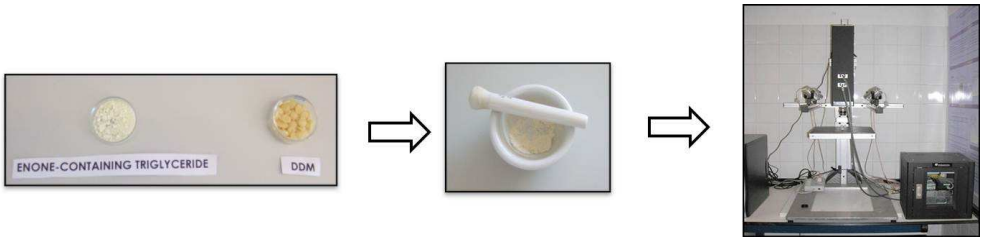


Figure 4. 1. Experimental procedure employed for the polymerisation process.

4.2. Results

Aza-Michael reaction with enone-modified vegetable oils. Evidence of the keto-enolic equilibrium by NIR chemical imaging and evolving factor analysis

Idoia Martí-Aluja, Itziar Ruisánchez, Virginia Cádiz, Santiago Maspoch and M. Soledad Larrechi

Analytical and Bioanalytical Chemistry 399 (2011) 1975 – 1982

Aza-Michael reaction with enone-modified vegetable oils. Evidence of the keto-enolic equilibrium by NIR chemical imaging and evolving factor analysis

Abstract. The existence of an enone-dienol tautomerism in enone-containing triglyceride, obtained from high oleic sunflower oil, was detected in an image set captured at 95ºC by Fixed Size Image Window-Evolving Factor Analysis (FSIW-EFA). A ^{1}H NMR spectrum and a UV-visible spectrum of the enone-containing triglyceride at 25ºC and at 95ºC were measured to corroborate the presence of the enol form. The presence of this equilibrium explains the different behaviour of the curing reaction between an enone-containing triglyceride and diaminodiphenylmethane, which was evaluated following the spectral evolution of two pixels that differ in the presence or absence of the enol form. As the enol form acts as inert specie in the reaction, it leads to different advance degree depending on which growing zone is observed.

Keywords – NIR chemical imaging; Crosslinking reaction; Enone-containing triglyceride; Fixed-size image window evolving factor analysis ; Chemometrics ; Tautomerism equilibrium

INTRODUCTION

Infrared Chemical Imaging that combines a Focal Plane Array detector with an Infrared spectrometer has been extensively used to investigate a wide variety of fields, ranging from polymers [1,2] to pharmaceutics [3-6], forensic science[7] and biological systems [8-11].

Near-Infrared Chemical Imaging (NIR Chemical Imaging) not only retains all the advantages provided by conventional Near-Infrared spectroscopy but also adds spatial information. This technique, in combination with the chemometric

analysis of the spectra recorded in all pixels of the image set has been extensively used to evaluate the inhomogeneity of final products [12-14] or to obtain quantitative information [15]. Nevertheless, the study of the spectral changes involved in a chemical reaction is mostly reported in the biochemist field [12], whereas in the analytical field it is not so usual. Specifically in the polymer field, where it is well known that the properties of the final cross-linked product depend not only on the experimental conditions but also on the pathway of the curing process [16]. In recent years, NIR Imaging technique has been referenced as a useful tool to evidence the presence of kinetic inhomogeneity in a curing reaction [17]. The present work illustrates an innovative application of NIR Imaging technique: the monitoring of the cure reaction between an enone-containing triglyceride, obtained by an environmentally friendly procedure from high oleic sunflower oil, and a diaminodiphenylmethane to obtain a polymeric network [18].

The reaction that takes place between them is an aza-Michael reaction, a variation in which the amine acts as the nucleophile and it has been extensively applied to the synthesis of crosslinked polymers derived from vegetable oils [19, 20]. In previous works, the complexity of that reaction was reported using conventional FT-IR and ^{1}H NMR [21,22], as well as the evidence of secondary reactions by NIR spectroscopy [23].

Fixed Size Image Window-Evolving Factor Analysis (FSIW-EFA) [24] was capable of detecting two significant factors in some pixels of the enone-containing triglyceride image set captured at the curing temperature. A component with high intensity values at the characteristic region of the O-H functional groups was detected in the enone-containing triglyceride image set by the local rank map obtained by FSIW-EFA. The presence of that increase in the intensity was attributed to the presence of an enone-dienol tautomerism, a chemical

equilibrium between a keto and an enol form. This fact was confirmed by ^{1}H RMN and UV-visible, showing that the tautomeric equilibrium is favoured to the keto form side. By the correlation coefficient map calculated between the characteristic enol spectrum and the average spectrum of the enone room temperature as a reference, it was possible to detect specific areas on the surface of the reaction image set where the concentration of the enol form was high enough to be detected. The evolution of the reaction at the characteristic absorption bands of the amine group was followed in two pixels, differing in the presence or absence of the enol form. Their different behaviour was attributed to the existence of the tautomeric equilibrium.

EXPERIMENTAL

Reagents

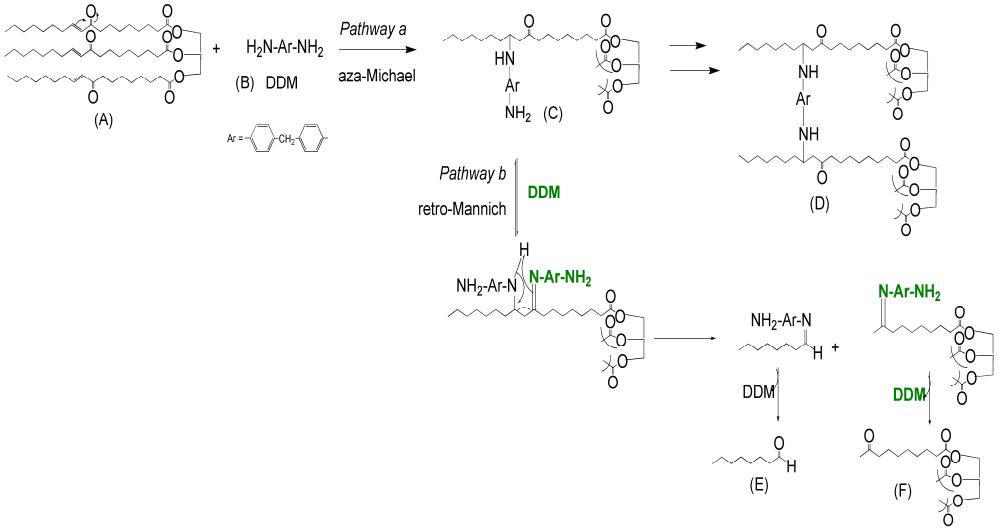
The enone-containing triglyceride (A) reagent was synthesized by the Polymer research group of the Universitat Rovira i Virgili and contains 2.6 α,β-unsaturated ketones per molecule [18]. The borontrifluoride monoethanolamine complex (BF_3·MEA) (Aldrich) and diaminodiphenylmethane (DDM) (Aldrich) were used as received.

Reaction conditions and procedure

The reaction that takes place between an enone-containing triglyceride (A), obtained from high oleic sunflower oil, and diaminodiphenylmethane (DDM) (B) is depicted in *Scheme 1, pathway a* [22, 23]. The primary amine group of reagent B reacts with the α,β-unsaturated ketone group of reagent A, to form the secondary amine (C) intermediate through an aza-Michael addition reaction, which evolves to finally obtain the desired polymeric network (D). However, there is a competitive reaction of this crosslinking reaction (*Scheme*

1, pathway b), where the condensation of the secondary amine (C) gives an intermediate, which suffers a retro-Mannich fragmentation and generates two subproducts, an aldehyde and a ketone (E and F).

The experimental procedure of the reaction involved mixing the necessary amounts of DDM and enone-containing triglyceride at room temperature to obtain a molar ratio of 1:1 (enone:DDM). The curing reaction was catalysed by BF_3·MEA, which was added to the mixture before heating. The mixture was uniformly deposited in a homemade circular crystal cell of 0.2mm thickness and a 1cm diameter and was introduced in an oven at 95ºC. This oven was located next to the Near-Infrared camera. The reaction was monitored for 120 minutes at 95ºC.



Scheme 1. Reaction between enone-containing triglyceride and DDM at 95ºC: Pathway a) aza-Michael reaction and Pathway b) retro-Mannich type fragmentation. (A) Enone-containing triglyceride, (B) DDM, (C) secondary amine intermediate, (D) polymeric network and (E, F) retro-Mannich fragmentation products.

Imaging acquisition and pre-processing

An image set of the pure reagents at room temperature (T=22ºC) and at 95 ºC were collected by a Near-Infrared Imaging spectrometer (Think Spectrally Roda-25, (Think Spectrally, Valencia, Spain)) with a mercury-cadmium-telluride (MCT) detector, providing a spatial resolution of 60 x 60 pixels. The size of each pixel was 100µm x 100µm and the spectral range was 1200 –2100 nm (or 8333-4762 cm^{-1}), every 7 nm, giving a total image-acquisition time of 120s. All image sets were transformed into MATLAB files in order to treat them [25].

An image set was captured every 5 minutes, obtaining 24 image sets during the reaction. To carry out the image acquisition, the crystal cell was moved quickly from the oven to the instrument sample holder, situated 13 cm below the objective. The site of the crystal cell was controlled to always capture the same position image set. While the sample was irradiated, the temperature of the crystal cell was controlled by a Temperature measuring probe, observing a variation between 95-90 ºC. It is possible to assume that this variation does not affect to the cure reaction.

Savitzky-Golay smoothing and Standard Normal Variate (SNV) were applied to all image sets in order to avoid undesirable phenomena commonly found in chemical images such as scatterings.

CHEMOMETRIC ANALYSIS

Fixed size image window-evolving factor analysis (FSIW-EFA)

The FSIW-EFA is a local rank exploratory method adapted to image analysis so that the neighbourhood around each pixel is preserved. This algorithm executes the Singular Value Decomposition (SVD) analyses in a fixed image

window or pixel area, which is moved one pixel at a time until the whole image is covered. In other words, the initial spatial structure of the image is preserved because the small windows are constructed by taking one pixel and all its surrounding neighbours. The singular values (SV) obtained in these local analyses are displayed together in singular value plots, where the spatial structure of the image is preserved. In these plots, large values are related to the presence of chemical constituents, which are significant contributions to the signal; whereas small values describe the experimental noise [26, 27].

Correlation coefficient maps

Before applying this methodology it is necessary to unfold the image set, which is crucial to adapt the 3-D array. Consequently, the two-dimensional structure **D** (MN x λ) is obtained when **D** (M x N x λ) is unfolded [28].

The correlation coefficient is a very useful tool to study the similarity between spectra. In chemical images, it is frequent to calculate the correlation coefficients between pure spectra of the analytes and each spectrum of each pixel of the image, which allow plotting a correlation map with the initial dimensions of the image. The correlation coefficient of pixels with spectra very close to the pure spectrum of an analyte is 1 or very close to 1, whereas a lower correlation coefficient (values close to 0) indicates that the similarity between the pixel and the analyte is practically null. The colour intensity of the correlation maps acts as a magnitude indicator, where hot colours refer to high correlation coefficients and cold colours are related to low correlation coefficients [29-31]. A routine called BACRA was used for the analysis of the correlation, its software provided by *Amigo J.M. et alt.* [32].

RESULTS AND DISCUSSION

The characteristic bands for the studied system are related to the amine groups present in the DDM reagent, which appear at 1508 nm (R_2NH and $R\text{-}NH_2$) and at 1984 nm ($R\text{-}NH_2$). In Figure 1, three images at 1984 nm recorded during the reaction are depicted: the first and the last reaction time images (Figs. 1a and 1c) and the image at t=60minutes, half way through the experimental cure reaction (Fig. 1b). The colour intensity is directly related to the intensity values measured, observing the different variation of the intensities between the first and the last image depending on which pixel is considered. It is well known that the cure processes are not homogeneous, so the non-homogeneity observed in the first image recorded at t=5minutes was analysed. According to the bibliography, at this time and in the experimental conditions, the vitreous state should not have been achieved, so the image ought to have been almost homogeneous. A possible explanation could be the non-uniform mixing of both reagents. To evaluate this possibility, an exploratory analysis of the images of the pure reagents captured at room temperature (22 ºC) and at 95ºC was done. In Fig. 2a the enone-containing triglyceride images at 1732nm (characteristic of the stretching vibration of its carbonyl groups) at both temperatures are represented. The histograms of each image were constructed in order to evaluate the intensity variations present in the previous figure, shown in Fig. 2b. The room temperature distribution can be regarded as that associated with experimental variability when a homogeneous sample is measured repeatedly (in the image set 3600 measurements are effectuated), whereas the 95ºC distribution presents a behaviour far from this situation. A symmetric distribution of the intensities was obtained at both temperatures analysing the DDM images and histograms obtained at 1984nm (characteristic of the primary amine groups present in the reagent (results not shown)). From this analysis, it

is possible to formulate a first hypothesis: the non-homogeneity observed at the beginning of the reaction (Fig. 1a) could be associated with the enone reagent behaviour, not with the cure reaction.

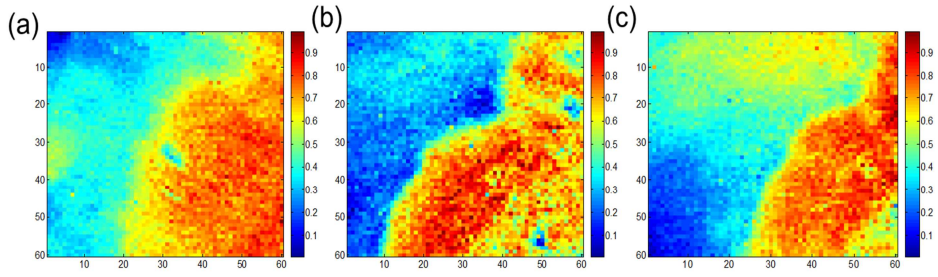


Figure 1. Images at 1984nm (normalised peak intensities) captured at 95ºC, during the cure reation: (a) t=5minutes, (b) t=60minutes and (c) t=120minutes.

To corroborate this statement, an individual analysis of each reagent at both temperatures was done by Fixed Size Image Window-Evolving Factor Analysis (FSIW-EFA), to assess the influence of the enone reagent heterogeneity in the reaction, and to obtain the reference experimental variability. A window size of (2x2) was used in order to preserve the spatial resolution of the image set as far as possible so that four singular values were obtained for each pixel.

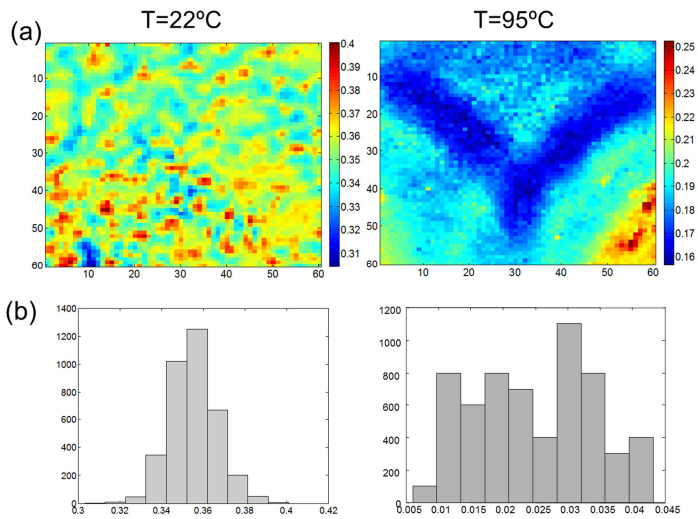


Figure 2. Images of the enone-containing triglyceride reagent at 1732nm captured at room temperature and 95ºC and (b) histogram plots of the images at 1732nm.

The first and the second singular value maps of the DDM and the enone reagent at 95ºC are represented in Fig. 3a and Fig. 3b, respectively. Larger values represent significant contributions to the signal related to the presence of chemical constituent, whereas smaller values describe the experimental noise. These plots give a quick qualitative idea of the zones where there is rank overlap.

Nevertheless, these representations are not very useful to select a numerical threshold value that separates the significant from the noise-related singular values. To achieve that, the representation based on the Parallel Coordinate System (PCS) [33-34] of the magnitude of singular values versus the four singular values was done and is represented in Fig. 3c and Fig. 3d. Each line in the PCS plot connects the singular values obtained in the SVD analysis of one pixel area, having as many lines as pixels (60 x 60). Fig. 3c clearly shows that only one component is detected in the DDM reagent image set. Considering the

maximum value of the second singular value for the DDM reagent (Fig. 3c) as the noise threshold value, in some pixels of the enone reagent image set the second singular value was significant (Fig. 3d). When the FSIW-EFA was performed in the enone reagent image set captured at room temperature, only one singular value was noteworthy on the whole surface, as well as for the DDM reagent. Thus, the presence of two singular values in some pixels of the enone reagent image set at 95ºC must be related to a variation in the reagent as a consequence of the temperature. The related local rank map of the enone reagent at 95ºC is depicted in Fig. 4a, which allows the regions where more than one contribution is present to be located.

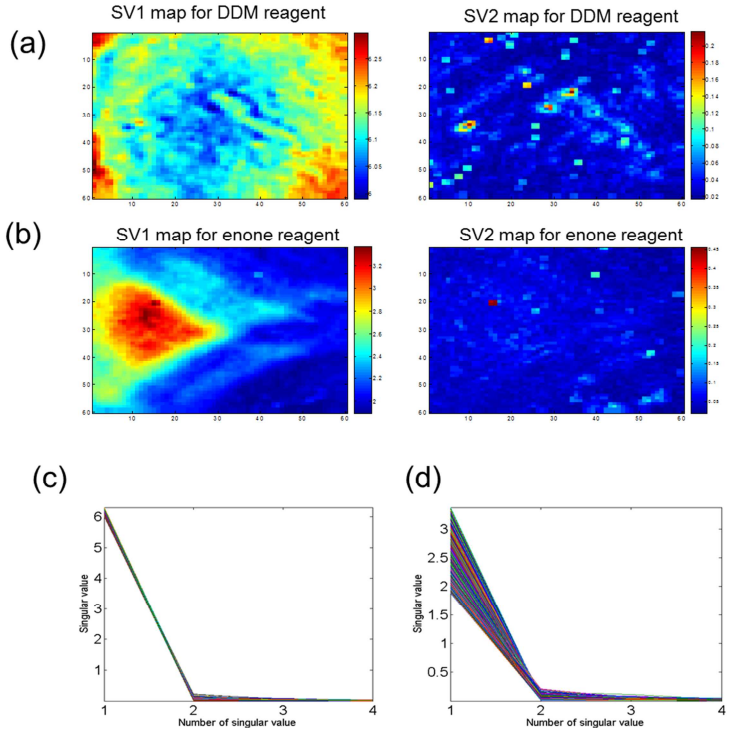


Figure 3. (a) First and second singular value maps for the DDM reagent image set at 95ºC, (b) first and second singular value maps for enone-containing triglyceride image set captured at 95ºC, (c) singular value plot for the DDM reagent and (d) singular value plot for enone-containing triglyceride reagent.

The average NIR spectrum of the enone reagent at room temperature (spectrum C) and the NIR spectrum obtained for the two pixels with a different rank value ticked in Fig. 4a (pixels A and B) are represented in Figure 4b. The pixel A spectrum differs significantly from the pixel C spectrum, its correlation coefficient being very low (R= 0.41). Moreover, the A spectrum presents high intensity values in the region between 1400-1600nm, a characteristic zone where the O-H group absorb. This functional group can be attributed to the presence of an enone-dienol tautomerism, a chemical equilibrium between a keto form and an enol form. Although it is not a favoured process, the presence of C-C double bonds in the enone derivative (-CH_2-CO-CH=CH-) can facilitate this equilibrium towards the dienol derivative (-HC=C(OH)-CH=CH-). To investigate that, the ^{1}H NMR spectra of the enone-containing triglyceride at room temperature and at 95ºC respectively were recorded (Figs. 5a and 5b). Comparing both spectra two new signals can be observed in the spectrum of the heated enone derivative, which appear at 3 ppm and 5.4 ppm. These peaks, as indicated in Fig. 5b, can be attributed to the H of the OH group and to the olephinic H of the new double bond of the dienol derivative. Moreover, the UV-visible spectra of the enone reagent at room temperature (solid line) and at 95ºC (dotted line) were recorded (Fig. 4c). At 95ºC the peaks are displaced to smaller wavelengths, a fact that is compatible with the presence of the enol form where a new double bond is formed.

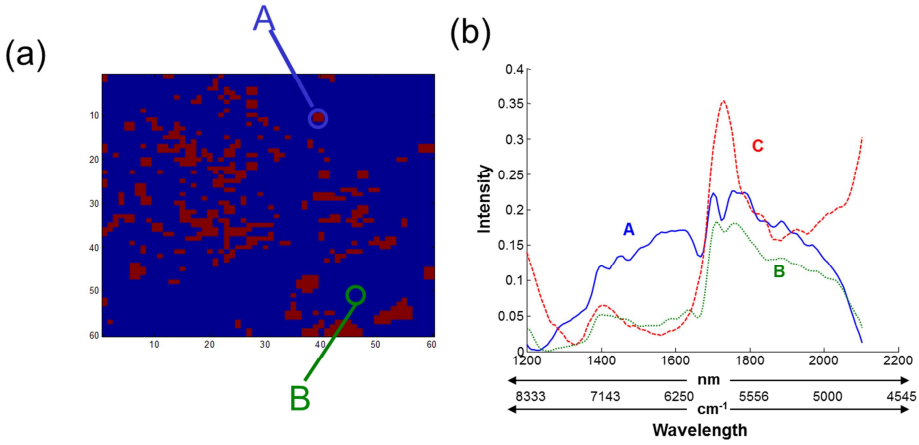


Figure 4. (a) Local rank map for the enone-containing triglyceride reagent image set captured at 95ºC and (b) spectrum of the ticked pixels for the enone reagent (A, solid line and B, dotted line) and average spectrum of the enone reagent image set captured at room temperature (C, dashed line).

At this point, the variability sources present in the enone reagent image set can be associated with a mixture of the species involved in the tautomerism. The characteristic spectrum of each specie is depicted in Fig. 4b: the A spectrum was associated with the enol form and the B spectrum was related to the keto form. The correlation coefficient between the B spectrum and the C spectrum is R=0.85. The two forms for the enone reagent were present at the experimental temperature, but their proportion was not the same. The keto form was greater and the enol form was not detected in all pixels, which is directly related to the singular values map represented in Fig. 3b.

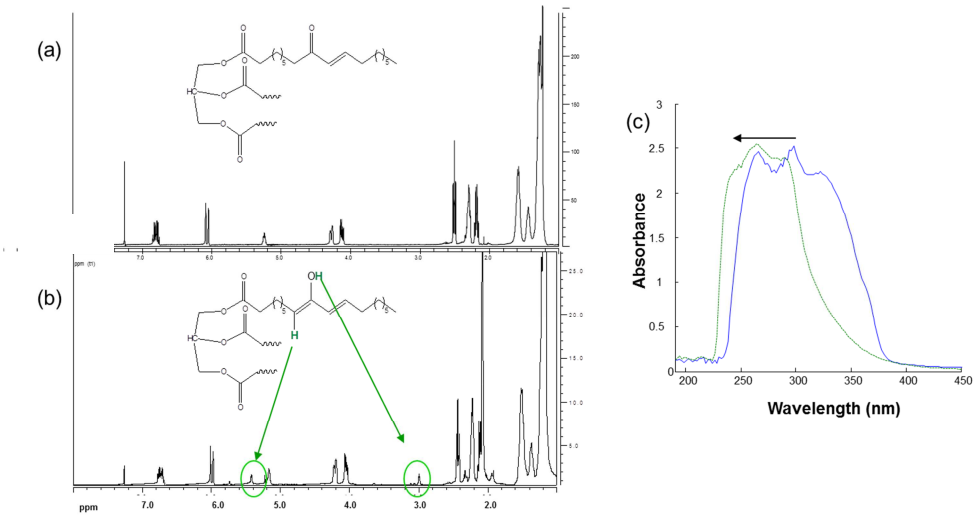


Figure 5. (a) ^{1}H NMR spectrum of the enone reagent at room temperature, (b) ^{1}H NMR spectrum of the enone reagent at 95ºC and (c) UV-visible spectra of the enone-reagent at room temperature (solid line) and at 95ºC (dashed line).

From the pure reagent analysis done and taking into account the non-reactivity of the enol with DDM, the initial experimental conditions could not be the same in all pixels recorded during the cure reaction, which contributes to the explanation of the non-homogeneity depicted in Fig. 1a. The enol form is not the predominant specie in the equilibrium and it is hidden in the image sets recorded. It is, therefore, necessary to find a characteristic pixel in the reaction image set where the equilibrium could be evidenced. To achieve that, the contribution of the DDM reagent at the first image was removed, subtracting the average spectrum of the unfolded image set captured with the equivalent amount of DDM reagent used in the mixture reaction. Then, the correlation coefficient between the recorded spectrum at each pixel of this new image set and the representative NIR spectrum of the enol form were calculated. The correlation coefficient map is represented in Fig. 6. As expected, the correlation coefficient values are small (R=0-0.25) as a consequence of the predominance of the keto form in all the pixels. The pixels where the enol form is favoured are

the ones with higher correlation coefficient values. Considering that, two pixels were selected to analyse the evolution of the reaction (ticked in Fig. 6): pixel C which was associated basically to the enone reagent, and pixel D, in which the proportion of the enol form is high enough to be detected. As the enol form does not react with the DDM reagent, at the beginning of the reaction the molar ratio between the two reagents will be higher in pixel C than in pixel D.

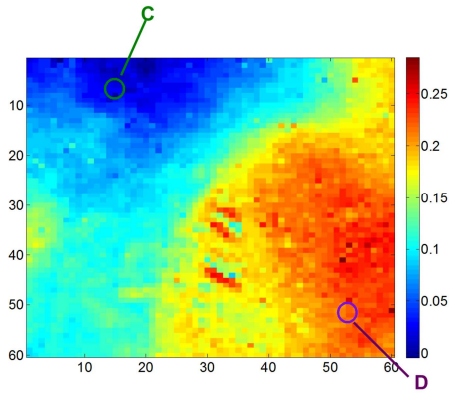


Figure 6. Correlation coefficient map between the image set at t=0minutes and the NIR spectrum of the enol form.

One way to analyse the effect of the different molar ratio along the image surface is to resolve quantitatively the concentration values for the chemical species involved in the reaction, considering the NIR spectrum of the two mentioned pixels throughout the reaction time. However, the presence of the enol form affects the kinetics of the reaction but not on its pathway. Therefore, a faster way to analyse this effect is to study the evolution of the intensity values throughout the reaction time at a characteristic absorption band of a reagent. The intensity values for pixel C (dotted line) and D (solid line) at 1984nm (characteristic of the primary amine group present in DDM reagent) are represented in Figure 7. The interpretation of the evolution of the reaction at 1984nm is not easy, because the aza-Michael reaction [23] is a complex

reaction that competes with a retro-Mannich fragmentation (*Scheme 1, pathway b*) in which DDM is newly formed. As a whole, during the first 15minutes of the reaction the most significant differences between both pixels can be observed, so the analysis was focused in that region. An increase in the intensity values is observed in pixel C, where the enol form is not enough to be detected. This increment can be associated with the existence of the retro-Mannich reaction, in which DDM is generated. On the other hand, there is a decrease in the intensity values during the first 15 minutes in pixel D, where the molar ratio between the reagents (enone:DDM) is lower than 1; so there is an excess of DDM. After this time, the behaviour of the amine band is similar in both cases, as was interpreted in previous works [18,22,23].

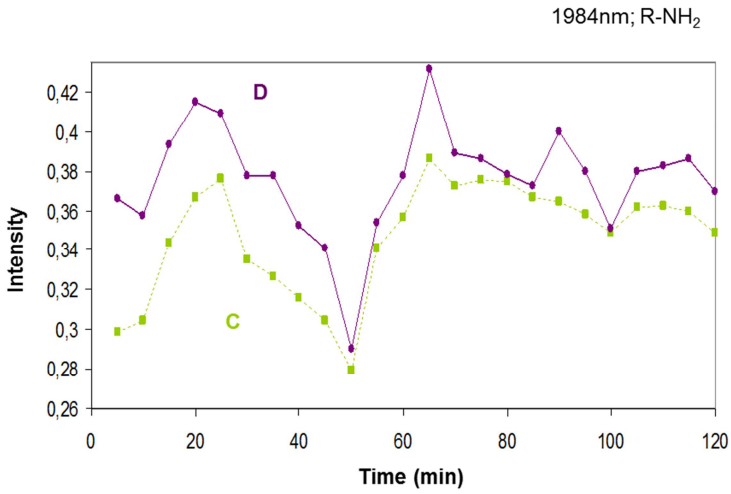


Figure 7. Intensity changes of the R-NH_2 band at 1984nm; when the enol form is present (pixel D, solid line) and the keto form is predominant (pixel C, dotted line).

CONCLUSIONS

NIR Chemical Imaging combined with the chemometric analysis has allowed evidencing an enone-dienol tautomerism when enone-containing triglyceride is heated. ^{1}H NMR and UV-visible spectroscopy have corroborated this fact. As a consequence of the reactions between enone-containing triglyceride with primary amines from the beginning, different reaction rates must take place depending on which growing zone is observed. From the obtained results we can conclude that better yields in this curing process will be obtained using a molar ratio enone: DDM higher than 1.

ACKNOWLEDGEMENTS. We would like to acknowledge the Spanish Ministry of Science and Innovation (Project CTQ2007-61474/BQU and Project CTQ2007-62528) for economic support and the Universitat Rovira i Virgili, for providing Idoia Martí-Aluja with a doctoral fellowship.

REFERENCES

[1] T. Ribar, R. Bargava, J. L. Koenig, *Macromolecules* **33** (2000) 8842 – 8849.

[2] S. J. Oh, J. L. Koenig, *Anal. Chem.* **70** (1998) 1768 – 1772.

[3] Y. Roggo, N. Jent, A. Edmond, P. Chalus, M. Ulmschneide, *Eur. J. Pharm. Biopharm.* **61** (2005) 100 – 110.

[4] C. A. Coutts-Lendon, N. A. Wright, E. V. Mieso, J. L. Koening, *J. Controlled Release* **93** (2003) 223 – 248.

[5] Y. Roggo, A. Edmond, P. Chalus, M. Ulmchneider, *Anal. Chim. Acta* **535** (2005) 79 – 87.

[6] K. L. A. Chan, S. G. Kazarian, *J. Comn. Chem.* **1** (2005) 301 – 305.

[7] J. R. Kalman, M. Tahtouh, *J. Forensic Sci.* **50** (2005) 1 – 9.

[8] J. Dubois, E. N. Lewis, F. S. Fry, E.M. Calve, *Food Microbiology* **22** (2005) 577 – 583.

[9] C. Krafft, M. Kirsch, C. Beleites, G. Schackert, R. Salzer, *Anal. Bioanal. Chem* **389** (2007) 1133 – 1142.

[10] Kohler, D. Bertrand, H. Martens, K. Hannesson, C. Kirschner, R. Ofstad, *Anal. Bioanal. Chem.* **389** (2007) 1143 – 1153.

[11] L. E. Nita, A. P. Chirac, I. Neamtu, M. Bercea, M. Pintilie, *Colloid Surface A* **348** (2009) 254 – 262.

[12] R. Gosselin, D. Rodrigue, C. Duchense*, Can. J. Chem. Eng* **86** (2008) 869 – 878.

[13] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, J*. Pharm. Biomed. Anal.* **44** (2007) 683 – 700.

[14] C. D. Ellison, B. J. Ennis, M. L. Hamad, R. C. Lyo*, J. Pharm. Biomed. Anal.* **48** (2008) 1 – 7.

[15] C. Gendrin, Y. Roggo, C. Collet, *Talanta* **73** (2007) 733 – 741.

[16] K. Dusêk, *Adv. Polym. Sci.* **78** (1986) 1 – 59.

[17] M. Fischer, C. D. Tran, *Anal. Chem.* **71** (1999) 953 – 959.

[18] L. Montero de Espinosa, J. C. Ronda, M. Galià, V. Cádiz, J*. Polym. Sci. A: Polym. Chem*. **46** (2008) 6843 – 6850.

[19] B. D. Mather, K. Viswanathan, K. M. Miller, T. E. Long, *Prog. Polym. Sci.* **31** (2006) 487 – 531.

[20] B. D. Mather, K. M. Miller, T. E. Long, *Macromol. Chem. Phys.* **207** (2006) 1324 – 1333.

[21] M. Galià, L. Montero de Espinosa, J. C. Ronda, G. Lligadas, V. Cádiz, *Eur. J. Lipid Sci. Technol.* **111** (2009) 1 – 10.

[22] L. Montero de Espinosa, J. C. Ronda, M. Galià, V. Cádiz*, J. Polym. Sci. A: Polym. Chem*. **48** (2010) 869 – 878.

[23] V. del Río, M. P. Callao, M. S. Larrechi, L. Montero de Espinosa, J. C. Ronda, V. Cádiz, *Anal. Chim. Acta* **649** (2009) 148 – 154.

[24] A. de Juan, M. Maeder, T. Hancewicz, L. Duponchel, R. Tauler, (2009) In: Salzer R, Siesler H W (eds) Infrared and Raman Spectroscopic Imaging, 1st edn WILEY-VCH, Weinheim .

[25] The Mathworks, MATLAB Version 7.0, Natick, MA, 2004.

[26] A. de Juan, M. Maeder, T. Hancewicz, R. Tauler, *Chemometr. Intell. Lab*. **77** (2005) 64 – 74.

[27] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, *Trends Anal. Chem.* **23** (2004) 70 – 79.

[28] P. Geladi, H. F. Grahn, J. E. Burger, (2007) In: Grahn H F, Geladi P (eds) Techniques and applications of Hyperspectral Image Analysis, 1st edn. WILEY-VCH, Chichester.

[29] J. M. Amigo, J. Cruz, M. Bautista, S. Maspoch, J. Coello, M. Blanco, *Trends Anal. Chem.* **27** (2008) 696 – 713.

[30] J. M. Amigo, C. Ravn*, Eur. J. Pharm. Sci.* **37** (2009) 76 – 82.

[31] C. Cairós, J. M. Amigo, R. Watt R, J. Coello, S. Maspoch, *Talanta* **79** (2009) 657 – 664.

[32] http://www.models.kvl.dk/users/jose_manuel_amigo/index.htm

[33] A. Inselberg, *Comput. Stat.* **14** (1999) 53 – 77.

[34] A. Inselberg, *Chemometr. Intell. Lab. Syst.* **60** (2002) 147 – 159.

4.3. General conclusions

NIR-CI provides valuable information about the chemical environment involved in the growth of a polymer. The fact that it can be used *in line* and its availability are two important features that mean that it can be regarded as a good alternative to more complex techniques.

The measurement of samples involves generating a data cube which contains a great deal of information. Chemometric techniques are needed to extract what is useful from this huge amount of data.

From an experimental point of view, temperature plays an important role in the quality of images. In the present study, we used a prototype in which NIR radiation impacted directly on the sample and caused temperatures of 50 ºC. It is important to bear this in mind when working with samples that can decompose at high temperatures.

In addition, I had the opportunity to work and compare images from two different optical lenses, acquired by the same instrument. The material from which the lens is made, affected the spectral resolution in the chemical images. The results of this comparative study are under preparation and they are not included in the thesis.

It is important to take into account both aspects – temperature and lens material – because the final conclusions extracted from the chemometrics point of view may be affected.

4.4. References

[1] H. R. Kricheldorf, O. Nuyken, G. Swift, *Handbook of polymer synthesis*, Marcel Dekker, New York, 2005.

[2] D. D. Andjelkovic, M. Valverde, P. Henna, F. Li, R. C. Larock, *Polymer* **46** (2005) 9674 – 9685.

[3] W. Kuhn, P. Barth, S. Hafner, G. Simon, H. Schneider, *Macromolecules* **27** (1994) 5773 – 5779.

[4] B. Blümich, P. Blümer, *NMR imaging of polymer materials,* Hüthig & Wepf, Basel, 1993.

[5] H. F. Grahn, P. Geladi, Techniques and applications of hyperspectral image analysis, WILEY-VCH, Chichester, 2007

[6] R. Salzer, H. W. Siesler, *Infrared and Raman spectroscopic imaging*, WILEY-VCH, Weinheim, 2009.

[7] J. M. Amigo, *Anal. Bioanal. Chem*. **398** (2010) 93 – 109.

[8] A. de Juan, R. Tauler, R. Dyson, C. Marcolli, M. Rault, M. Maeder, *Trends Anal. Chem*. **23** (2004) 70 – 79.

[9] P. L. Nayak, *Polym. Rev*. **40** (2000) 1 – 21.

[10] K. Perepelkin*, Fibre Chem.* **37** (2005) 417 – 430.

[11] V. Sharma and P. P. Kundu, *Prog. Polym. Sci*. **33** (2008) 1199 – 1215.

[12] A. A. Ardakani, J. D. Gelorme, L. L. Kosbar, *US Patent* **5** (1998) 833 – 883.

[13] M. Galià, L. Montero de Espinosa, J. C. Ronda, G. Lligadas, V. Cádiz, *Eur. J. Lipid. Sci. Tech.* **112** (2010) 87 – 96.

[14] J. D. Earls, J. E. White, L. C. López, Z. Lysenko, M. L. Dettloff, M. J Null, *Polymer* **48** (2007) 712 – 719.

[15] L. Montero de Espinosa, J. C. Ronda, M. Galià, V. Cádiz, *J. Polym. Sci. A: Polym. Chem.* **48** (2010) 869 – 878.

[16] V. del Río, M. P. Callao, M. S. Larrechi, L. Montero de Espinosa, J. C. Ronda, V. Cádiz, *Anal. Chim. Acta* **649** (2009) 148 – 154.

Chapter 5

Analysis of aggregation process by FTIR spectroscopy

5. 1. Introduction

The overall objective of Chapter 5 is to develop fast analytical methods based on the chemometric analysis of infrared spectra recorded during insulin aggregation, and to analyse the importance of this process. Insulin is one of many proteins that undergo aggregation and that is often used as a model [1-3]. Many spectroscopic and microscopic techniques have been used to study aggregation processes, including fluorescence spectroscopy [4-8], circular dichroism (CD) [9-13], Raman spectroscopy [14-17] and scanning electron microscopy (SEM) [18,19]. In the present thesis, infrared spectroscopy was chosen to monitor insulin aggregation because it is simple, fast and available in most laboratories. A deeper explanation of previous studies is described in the introduction of the published works, included in the results section.

Further, the features of insulin aggregation are listed. Insulin is a small protein consisting of two polypeptide chains: the A chain (21 aminoacids) and the B chain (30 aminoacids), linked by two interchain disulphide bridges [1-3, 20, 21]. The polypeptide chains adopt a helicoidal conformation that forms the secondary structure called α-helix [22,23]. Insulin is synthesized in the pancreas as hexamer, an inactive form that is very stable and consists of the union of six monomer units coordinated by Zn [4, 20, 24-31]. Nevertheless, the active form of this protein is the monomeric (a single insulin molecule) which is less stable because of its tendency to associate into dimers by means of hydrogen bonding. The monomeric forms of insulin have a natural tendency to self-assemble to form dimers, tetramers and hexamers which are stabilised by hydrophobic interactions and non-covalent bonds. In some conditions, these assembled forms can precipitate to form spherical arrangements with the same native structure of insulin. This natural process is called association and a schematic representation is shown in Scheme 5.1.a) [32-36]. The monomeric

insulin can also unfold, thus losing the initial α-helix structure to form aggregates. Although the mechanism of aggregation is still under review, it has commonly been described as is shown in Scheme 5.1.b). In this process, there is an initial unfolding of the monomeric forms that lead to unfolded species with a high content of random coil, β-turns and other unordered secondary structures. Once there are a significant number of unfolded forms, they associate to form nuclei. These species of nuclei are mainly composed of β-sheet structures, which bind together by means of covalent bonds to form insoluble aggregates (only β-sheet structures). In this process the native structure of insulin (α-helix) is completely converted to β-sheet structures [37-44].

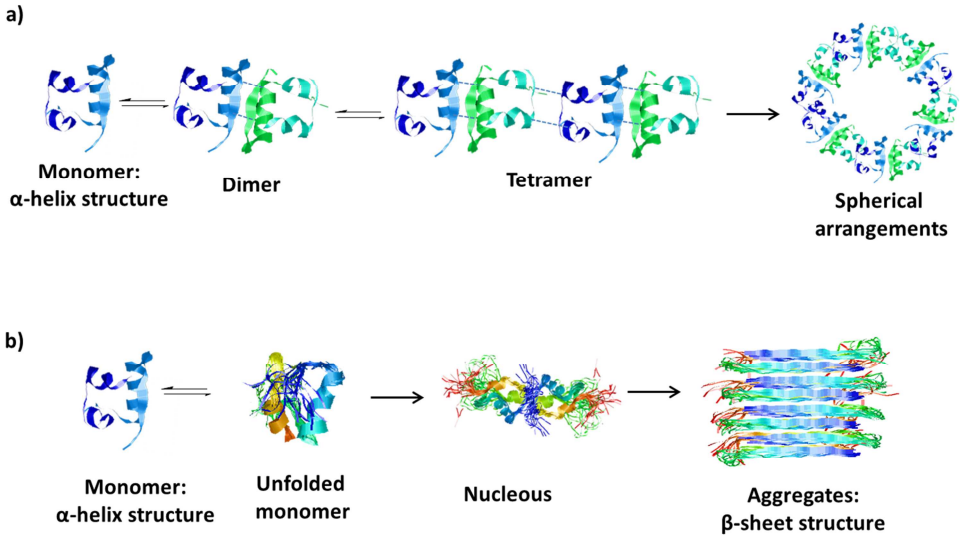
Overall, protein aggregation gives rise to the formation of insoluble forms with an alteration in the native structure of proteins. Understanding the mechanism of protein aggregation and the variables that influence the process is currently the focus of various research areas [28, 31, 41]. In medicine, it is of great interest because the abnormal formation of protein aggregates has been associated with such diseases as Alzheimer's, Parkinson's, Huntington's or Creutzfeldt-Jakob diseases [42, 44, 45]. The aggregation process destabilises native proteins, which also concerns the pharmaceutical industry because it is essential to maintain the stability of proteins during their production and storage as pharmaceutical products [47,48].

This chapter has two main objectives:

- To assess the influence and effect of antiretroviral drugs on the aggregation process of insulin (section 5.2.).

- To go deeper into the natural processes that can occur in insulin, studying the influence of biochemical variables and finding the

response surface of the relationship between the aggregation time and the values of the biochemical variables (section 5.3.).



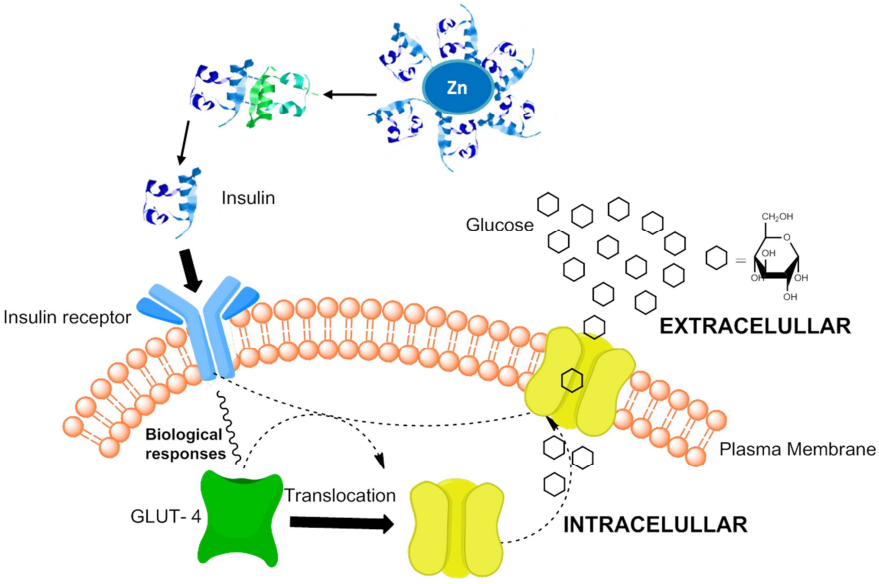
Scheme 5. 1. Schematic representation of a) insulin association and b) insulin aggregation.

Each objective is introduced and explained before the results obtained are discussed. At the end of this chapter, the general conclusions drawn from the results are included.

5. 2. Insulin aggregation induced by antiretroviral drugs

Prolonged treatment of antiretroviral drugs in acquired immunodeficiency syndrome (AIDS) patients commonly causes such metabolic disorders as insulin resistance [49-52]. Insulin resistance is a disorder that occurs when normal levels of insulin are insufficient to stimulate glucose uptake in the insulin signalling pathway [52-55]. Several theories relate this disorder with the

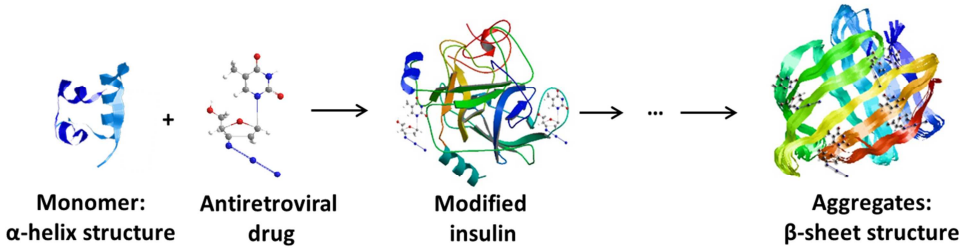
alteration of some of the elements involved in the glucose metabolism [54,55]. One suggests that the structure of insulin is altered, which disables the linkage to the insulin receptor. A schematic representation of the process that insulin undergoes during the glucose metabolism is represented in Scheme 5. 2. First, there is an internal conversion from the hexameric to the monomeric form [56-57]. These active forms bind to the insulin receptor located outside the cell membrane. When this linkage takes place, some biochemical and biological stimuli occur that induce the translocation of glucose transporter type 4 (GLUT-4) to the plasma membranes. This translocation facilitates the entry of glucose into the cell [54, 58-59].



Scheme 5. 2. Schematic representation of insulin role in glucose metabolism.

Some antiretroviral drugs contain functional groups that are capable of reacting with the aminoacids present in insulin [60,61]. Therefore, they can modify the structure of insulin and prompt it to unfold, which induces the aggregation [62,63]. Scheme 5.3. shows a schematic representation of this process. As a

result, the native structure of insulin is modified, which may produce an alteration in the normal function shown in Scheme 5.2.



Scheme 5. 3. Scheme of the alteration of insulin in presence of antiretroviral drugs.

The work developed on assessing the effect of antiretroviral drugs on insulin aggregation was carried out in two stages. First, an analytical method based on the principal component analysis (PCA) of infrared spectra was established to find an appropriate response variable that could quantify the effect of an antiretroviral drug on the aggregation. The chosen drug was 3'-azido-3'-deoxythymidine (AZT) because it is one of the most widely used drugs for treating AIDS. In this step, the effect of the concentration of AZT on insulin aggregation was evaluated. The influence of serum proteins on the phenomenon was assessed using human serum.

The second stage of the work was focused on the quantitative analysis of the effect of three different types of antiretroviral drugs on insulin aggregation. Zidovudine (AZT), ritonavir and efavirenz were the three drugs selected. They belong to different classes of antiretroviral drugs commonly used in highly active antiretroviral therapy (HAART) [64-67]. The analytical method was based on the multivariate curve resolution-alternating least squares (MCR-ALS) of the infrared spectra recorded during the process. This method provided

information about the influence of antiretroviral drugs on the mechanism and pathway of aggregation.

The results section contains detailed information about the experimental procedure used in both cases as two research papers.

5.2.1. Results

Chemometrics analysis of insulin aggregation induced by an antiretroviral drug (AZT)

Idoia Martí-Aluja, Joan Ferré, Itziar Ruisánchez and M. Soledad Larrechi

Chemometrics and intelligent laboratory systems 118 (2012) 180 – 186

Chemometrics analysis of insulin aggregation induced by an antiretroviral drug (AZT)

Abstract. The analysis of the effect of drugs in insulin aggregation is interesting due to its influence on human metabolism. We present a method to analyse the effect of an antiretroviral drug (AZT) on the insulin aggregation, based on the principal component analysis of the infrared spectra recorded in the process. This method allows the analysis of the conformational changes involved in the aggregation, providing a time value associated with the aggregates formation. The analysis was performed using Zn-free insulin in buffered solution at 1.72 mM concentration. The aggregation was monitored *in situ* by FTIR at 37ºC and the dependence of the concentration of AZT was considered, contemplating three concentration levels of AZT. The interaction between AZT and insulin was proved by ^{1}H NMR spectroscopy. The aggregation states were characterised by solubility tests and microscope imaging. A comparative analysis of time aggregation led observing that the process was delayed when the concentration of AZT was lower than 1.72 mM, while it accelerated at equal or higher concentrations. This dependence was related to the concentration of the monomeric species in the medium. The analysis was performed in serum to evaluate the influence of serum proteins in the phenomena, concluding that AZT binds to insulin before than to serum proteins.

Keywords – Insulin; Insulin aggregation; AZT/Zidovudine; Infrared spectroscopy (FTIR-ATR); Principal Components Analysis (PCA)

INTRODUCTION

Insulin is a protein that has a key role in the glucose metabolism. Although insulin functions in the bloodstream take place as a Zn-free monomer, it can assemble into dimers or hexamers, the latter in presence of Zn [1-5]. The aggregation of insulin monomers to form oligomers is an intriguing and widely

studied phenomenon [6-9]. This process takes place due to the natural propensity of insulin to self-assemble and it is affected by, among others, temperature [3,5,10-12], acidity [11-13], ionic strength [5,13,14] and the presence of other molecules [15,16].

AIDS (Acquired Immune Deficiency Syndrome) patients under a prolonged treatment of AZT (Zidovudine, 3'-azido-3'-deoxy-thymidin) frequently develop insulin resistance, a physiological condition where the natural insulin become less effective [17,18]. The presence of drugs in the bloodstream may be related with this disease as long as drugs are able to modify the natural aggregation of insulin.

Bioinformatics tools applied to laborious experimental data have proved useful for timely providing deep insights on the drugs behaviour in biological medium both in basic research and drug development [19-23]. In this context, the aim of the present work was to conduct, for the first time, a chemometrics analysis of the insulin aggregation process induced by the antiretroviral drug AZT.

To investigate the insulin aggregation, many spectroscopic techniques have been employed, including circular dicroism (CD) [24,25], fluorescence [26,27], ^{1}H NMR [28,29] and Raman spectroscopy [30,31]. Infrared (IR) spectroscopy is commonly employed in the analysis of the structural changes of proteins [32-34]. This technique easily allows the monitoring *in situ* of the aggregation, so that has been selected in the present work. In this region, the amide I band is a sensitive marker of peptide secondary structure, as the vibrational frequency of each C=O bond depends on hydrogen bonding and the interactions between the amide units, both of which are influenced by the secondary structure.

The analytical methodology proposed in this work is based on the Principal component Analysis (PCA) of the recorded FTIR spectra. Chemometrics

methods based on factor analysis have been proved to be powerful techniques for the study of other protein processes [35-38]. Among them, PCA is a useful tool to evidence physicochemical transitions during conformation changes of proteins and other biological complex systems [39-43]. The insulin aggregation induced by AZT is studied for the first time using both techniques.

The analysis was performed using Zn-free insulin, initially in aqueous solution at 1.72mM concentration. Then, the analysis was carried out in a serum matrix to mimic the human body conditions and to evaluate the competitiveness between insulin and serum proteins.

The aggregation was monitored *in situ* by FTIR at 37ºC and the dependence of the concentration of AZT was considered, so that three experimental runs at different insulin:AZT molar ratios (1:0.25, 1:1 and 1:2) were performed.

The time-dependent spectral changes in the amide I region, characteristic of the secondary structure of proteins, were analysed by Principal Component Analysis (PCA). The scores plots were representative of a two state process. The aggregates were characterised by solubility tests and microscope imaging. PCA analysis provided a time value between both stages, useful to analyse the AZT effect on the insulin aggregation. In absence of AZT, a unique transition time at 38 minutes was detected.

The interaction between AZT and insulin, proved by ^{1}H NMR spectroscopy, and the analysis of the effect of AZT in the studied phenomenon, shows that depends on its concentration. The aggregation was delayed at a concentration of AZT lower than 1.72 mM, whereas at an equal or higher concentration, this process went faster and throughout the formation of soluble or/and insoluble aggregates depending on the concentration of AZT. The analysis was also

performed in a serum matrix to evaluate the influence of serum proteins in the process, concluding that AZT binds to insulin before than to serum proteins.

In our opinion, the biological significance of the results not only resides in contributing and understanding causes involved in insulin resistance observed after the prolonged treatment with this drug, which is actually a challenge, but also in controlling the intake of this drug in the human metabolism cycle.

MATERIALS AND METHODS

Chemical reagents

Zn-free human insulin in a HEPES sodium salt buffer (pH=8.2) at a concentration of (1,72mM) was purchased from Sigma-Aldrich (St. Quentin Fallavier, France) and used without further purification. The pharmaceutical product 3'-azido-3'-deoxythymidine (Zidovudine, AZT) was purchased in a commercial chemist. A human serum type AB was acquired from Sigma-Aldrich at stored at -20ºC. All reagents were used as received.

Experimental procedure and FTIR acquisition

The aggregation of native insulin was studied first at a concentration of 1.72 mM. This concentration value was chosen because it is within the interval of concentrations used in other aggregation studies. Secondly, insulin was mixed with AZT and the aggregation of this "altered insulin" was studied. Two sets of three experimental runs at different insulin:AZT molar ratios (1:0.25, 1:1 and 1:2) in absence or presence of human serum were monitored. In addition, three control experiments with the same AZT quantity in human serum were carried out. The experimental procedure involved mixing the necessary amounts of

insulin and AZT. Immediately, a drop of the mixture was placed on a small diamond crystal in the spectrophotometer ATR cell (FTIR 680 Plus JASCO and a RS232 Control) to take the measurements, which it was continuously purged with N_2 during the FTIR analysis. The FTIR spectra were acquired *in situ* every minute during 45 minutes, in the spectral range 1550-1750 cm^{-1}. The CO_2 contribution was removed with the control software Spectra Manager before exporting the spectra into Matlab [44]. All experiments were carried out at 37ºC, to mimic the human body conditions.

Chemometric analysis

The recorded infrared spectra from experiment *n* were arranged as rows in matrices $\mathbf{D_n}$ of size 45 x 209. Firstly, Savitzky-Golay smoothing was applied to all data sets to suppress the instrumental noise. Then, spectra were normalized to have the same area. PCA of the mean-centred spectral data was used to reveal spectral trends over time. PCA is a multivariate representation technique that decomposes the spectral data matrix as the product of two reduced-size matrices: a scores matrix, which corresponds to a compressed view of the spectra, and a matrix of loadings that contains the contribution of every original variable in the compressed data. The scores and loadings plots were used to detect groups of the spectra recorded throughout time and to relate them with the spectral bands and the involved changes.

Auxiliary instrumental techniques

^{1}H NMR spectra were performed in a Varian Gemini 300 NMR spectrometer at 75.4 MHz operating with proton noise decoupling.

The optical images were collected using a Leica DM 2500 confocal microscope, with a 100W illumination system and using a 10x objective.

RESULTS

Native insulin aggregation analysis

Infrared normalized spectra of insulin (1.72mM) acquired *in situ* over 45 minutes are presented in Figure 1. Although in Zn-free buffered solution insulin exists mainly as monomer [2,5], its propensity to self-associate leads to a small percentage of dimer species in equilibrium with the monomer [31]. The coexistence of both forms is revealed by the 1636 cm^{-1} band, that is indicative of the amide I marker for the β-sheet modes in the dimer, and the band at 1651 cm^{-1}, that corresponds to the α-helical structure present in both the monomeric and dimeric forms [34].Regarding the spectral evolution, a shift of the maximum absorbance from 1636 cm^{-1} to 1651 cm^{-1} suggests a partial loss of the initial β forms of insulin throughout time at the experimental temperature.

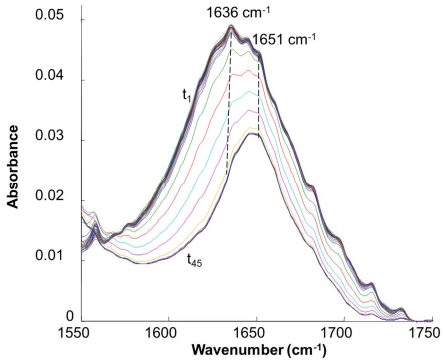


Figure 1. Infrared spectra of the native insulin obtained every minute at 37ºC, during 45 minutes.

Insulin aggregation induced by the self-association of the native protein is commonly described in the literature. The pathway of this process can be represented as:

Native insulin ⇌ Partial unfolding → Aggregates

where the formation of aggregates requires the partial unfolding of native insulin, which is responsible of the irreversibility of this process. The extension of this process is proportional to the concentration of monomers in the native protein [7].

According to the literature, some spectral changes related to the aggregates formation are expected, specifically the transition from a α-helix (1651 cm^{-1}) to a β-sheet structure (1625 cm^{-1} and 1680 cm^{-1}) [46-49]. Although these spectral changes were not observed in Figure 1, the decrease in intensity of the absorption bands over time may suggest that the molecules are vibrating in a more rigid medium, as it happens when aggregates are growing.

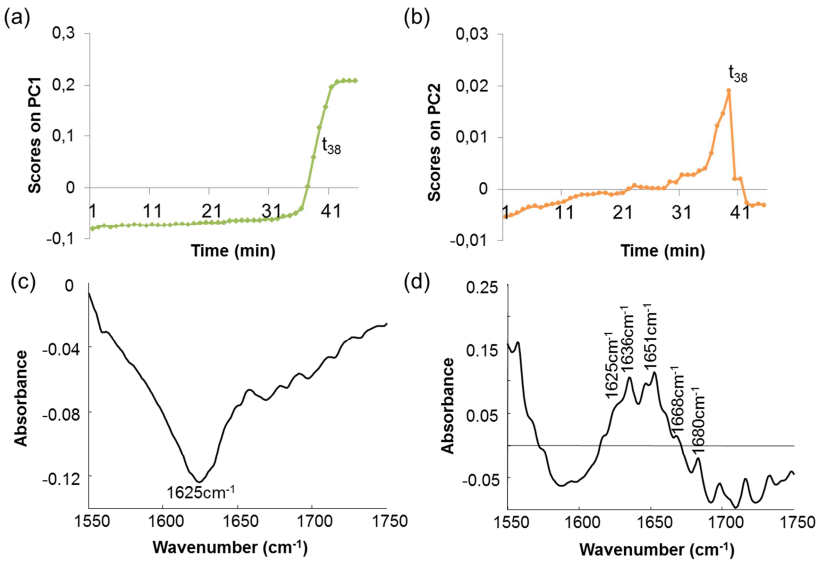


Figure 2. Scores plot on the (a) first Principal Component (PC) and (b) second PC and loadings plot on the (c) first PC and (d) second PC for the native insulin experiment. PC1 and PC2 respectively account for 98% and 2% of the data variability.

A better evidence of the changes in the infrared spectra throughout time was obtained by PCA. Figure 2a shows the scores plot for the first Principal component (PC), which accounted for the 98% of the data variability. The scores show an abrupt separation between the spectra recorded before 38 minutes and the ones collected subsequently. The physic-chemical interpretation can be obtained from the PCA loadings. The loadings plot for the first PC (Fig. 2c) shows the large importance of a band around 1625 cm^{-1}, associated with the intermolecular β-sheet structures, a sign of insulin aggregates [48,49]. Thus, the first PC shows that the main variations in the spectra were related to the changes involved in the conversion of insulin soluble α-helical forms into insoluble aggregates, enriched with β-structures [46-50].

Complementary details of Zn-free insulin are observed in the scores plot for the second PC (Fig. 2b). The scores steadily increase until the crossing point at 38 minutes, which coincides with the time observed in the first PC that was attributed to the insulin aggregation. The loadings of the second PC (Fig. 2d) show an important band at 1636 cm^{-1}, attributed to the initial β-structures of the dimeric forms, and the band at 1651 cm^{-1} related to the α-helical structure in the monomer. Thus, this second PC can be considered as representative of the dimer dissociation process, where new monomers are being released in the medium. The low percentage of spectral variability accounted by this PC (2%) was not surprising, considering the small concentration of the dimer species in the analysed Zn-free insulin. Less intense contributions in the loadings for the second PC (Fig. 2d) were at 1625 cm^{-1}, 1668 cm^{-1} and 1680 cm^{-1}, that are related to intermolecular β-sheet, β-turn and β-sheet structures respectively, characteristic of the formation of aggregates. Thus, the dimer dissociation and

insulin aggregates formation were taking place from the beginning, in agreement with reference [34].

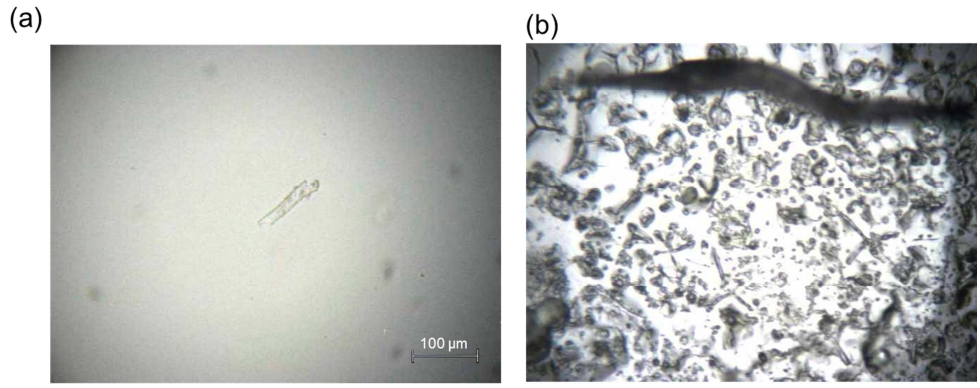


Figure 3. Microscopic images of insulin (1.72mM) captured at 37ºC (a), during 45 minutes: (a) t=0minutes and (b) t=45minutes.

These results suggest that insulin naturally self-assembles at 37ºC, producing a displacement of the dimer-monomer equilibrium toward the monomeric form. The crossing point at 38 minutes corresponds to the moment when the concentration of the unfolded monomer was high enough so that the formation of aggregates took place in a sufficient extend to be detected by infrared spectroscopy. Imaging by an optical microscope verifies the above observations (Fig. 3). No sign of the distinctive and characteristic structure of the aggregates is observed in the first image (Fig. 3a). In the second image (45minutes, Fig. 3b), a network of aggregates emergence and clusters of aggregates are present. Additionally, the presence of insulin aggregates was confirmed by a solubility test at 38 minutes. The final product was not only insoluble in polar solvents (propanol, methanol and ethanol) but also in non-polar solvents (THF, hexane and petroleum ether).

Insulin aggregation induced by AZT

The final amino groups of insulin are capable to open the cyclic imide structure present in AZT, forming the "AZT-altered insulin" compound which contains new amide / peptide bonds. This statement is in agreement with the ^{1}H NMR analysis. Figure 4 shows the ^{1}H NMR spectra of native insulin (1.72mM), AZT and the mixture of both at 1:1 molar ratio. For clarity, the spectra have shifted up in the y-axis direction and only the amide region between 6 and 8 ppm is shown. New bands in the amide region are present in the ^{1}H NMR spectra of the AZT-insulin mixture comparing to the native insulin [28,29], evidencing the interaction between both compounds. The extension of the interaction is related to the amount of AZT in the medium, so three experiments at different AZT concentration levels were carried out.

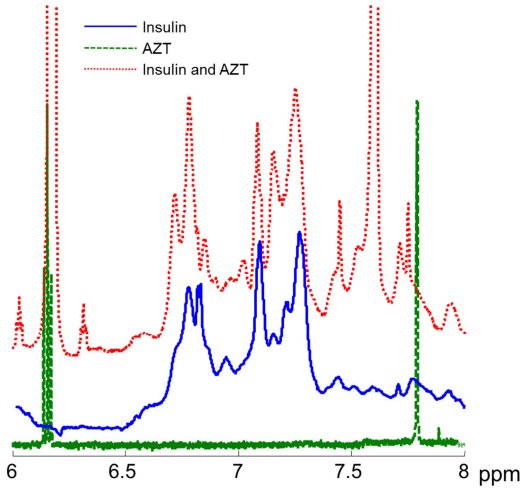


Figure 4. ^{1}H NMR spectra of native insulin (solid line), AZT (dashed line) and AZT-insulin mixture (dotted line).

In the first experiment, AZT was at a concentration lower than the native insulin concentration (1:0.25). In this case, no significant spectral variations were

observed over time (Fig. 5a), which indicates that the formation of aggregates did not take place. The microscopic images captured initially and at 45 minutes validate this statement due to its similarity to the image captured in Figure 3a (results not shown. Taking into account that the extension of the insulin aggregation is proportional to the monomer concentration [8], a plausible explanation is that the monomer concentration decreased as the amount of "AZT-altered insulin" complex increased, so that the formation of aggregates does not take place.

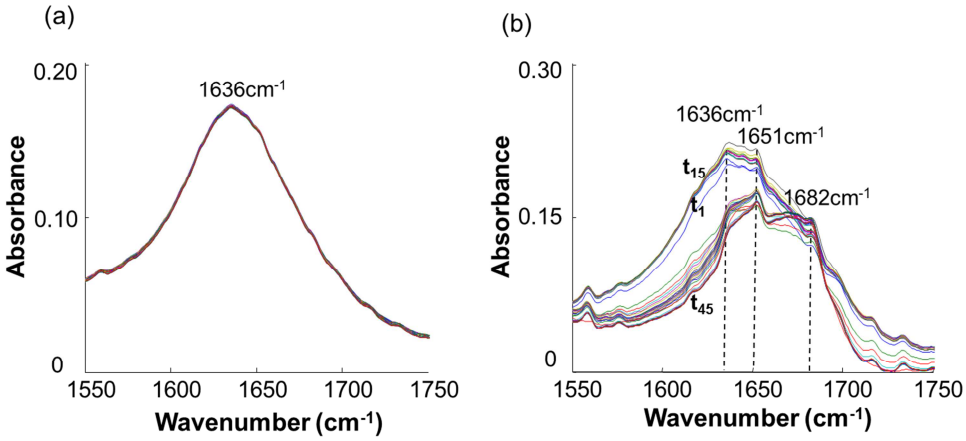


Figure 5. Infrared spectra of insulin: AZT obtained every minute at 37ºC, during 45 minutes for (a) the first experiment (1:0.25 molar ratio) and (b) second experiment (1:1 molar ratio).

In the second experiment, AZT had the same stoichiometric ratio as insulin (1:1). In this case, spectral features characteristic of the conformational changes involved in the process were identified (Fig. 5b). An increase in all absorption bands was observed during the first 15 minutes, which indicated new intermolecular interactions in the medium. After this time, not only was detected a shift on the maximum of the first spectrum (1636 cm^{-1}) to higher wavenumbers (1651 cm^{-1}) but also emerged the spectral band at 1682 cm^{-1},

which is characteristic of the β-aggregates resulting from the intermolecular folding of β-sheets [48,49]. The microscopic images recorded in that case (Fig. 6a and 6b) also verify the above discussion. In the third experiment, where the AZT concentration was twice the concentration of insulin (1:2), the spectral changes were similar but they took place faster (results not shown).

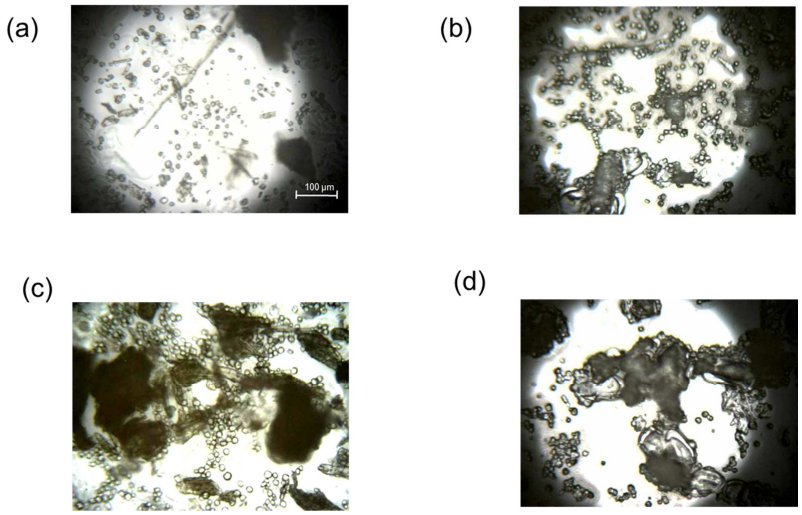


Figure 6. Microscopy images of insulin: AZT captured at 37ºC, during 45 minutes for the second experiment (1:1) at (a) t=0minutes and (b) t=45minutes; and for the third experiment (1:2) at (c) t=0minutes and (d) t=45minutes.

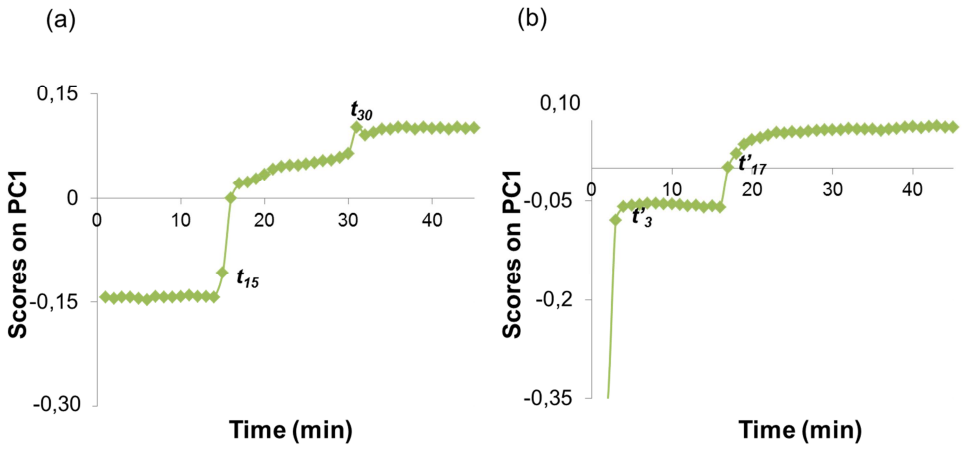


Figure 7. Scores plot on the first PC of insulin: AZT at (a) 1:1 and (b) 1:2 molar ratios. PC1 accounts for 99.0 % of the data variability in (a) and 99.8% in (b).

In the third experiment, the AZT concentration was twice the concentration of insulin (1:2). In this case, the observed spectral changes were similar than those in the second experiment but they took place faster (results not shown).

The second and third experiments were further studied by PCA. PCA of the first experiment was unnecessary due to the limited spectral variation over time.

This lasts experiments were further studied by PCA. For the second experiment, Figure 7a shows the scores on the first PC that represent the changes on the "AZT-altered insulin" over time. Compared with the scores of Zn-free insulin, in this case, two transition points are observed at 15 minutes and 30 minutes, together with a gradual evolution of the scores between both midpoints. The presence of the two transition points indicates that the aggregates formation came about in two stages. In the first stage, the opening of the cyclic amide group generates new alcohol groups (-OH) in the "AZT-altered insulin" structure. This fact is corroborated by the fact that the sample taken at minute 15 was soluble only in polar solvents. These alcohol groups may have a relevant

role in the aggregation pathway due to new intermolecular interactions, through H-bonding. In addition, the presence of the new intermolecular bonds can also explain the increase on the absorption bands during the first 15 minutes seen in Figure 5b. In the second stage, the soluble oligomers turn into insoluble aggregates [8]. This was confirmed by a solubility test after 30 minutes, which showed that the sample was insoluble in polar and non-polar solvents. Hence, a plausible mechanism of the "AZT-altered insulin" aggregation goes through the formation of soluble aggregates and can be expressed as:

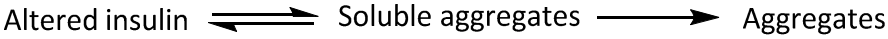
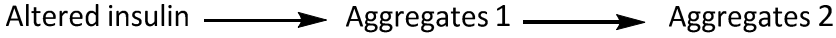


Figure 7b shows the scores on the first PC for the third experiment, at the highest AZT concentration. Again two transition points are observed but earlier and the scores profile did not show any spectral variation between both points. This can be explained by the earlier formation of insoluble aggregates due to the fact that the excess of AZT was incorporated in the aggregates network from the beginning [7,9]. Solubility tests at 3 minutes and 17 minutes showed that the samples were totally insoluble in polar and non-polar solvents. This test indicated that when AZT concentration is higher than insulin, insoluble aggregates are being formed in the initial minutes of the process and the following pathway can be proposed:



The first image captured (Fig. 6c) reveals that aggregates are growing from the very beginning, in agreement with the earlier formation of insoluble aggregates (results above). In the second image (45minutes, Fig. 6d), clusters of aggregates

are present resulting in a more rigid and compact network than in experiment 1:1 (Fig. 6b).

The infrared results together with the captured images corroborate the presence of different insulin aggregates at both molar ratios, so that the effect of the AZT concentration in this process.

Insulin aggregation induced by AZT in physiological serum

The analysis was also performed in presence of serum to evaluate if the previously behaviour observed also occurs when serum proteins can interact with AZT. The experiments were carried out at the three same molar ratios as previously, but only the results from the 1:1 ratio experiment will be discussed. Then, the interaction between AZT and serum proteins was analysed by infrared spectroscopy (control experiment, Figure 8a). Taking into account the complexity of serum proteins, more absorption peaks in the amide I region emerged from the beginning [50]. Focusing the attention on the absorption bands at 1636 cm^{-1}, 1651 cm^{-1} and 1681 cm^{-1}, its behaviour was similar to the one already commented in Figure 3b except that a steady decrease in the intensity bands over time was observed, which probably indicated the self-association of serum proteins. When insulin was present (Fig. 8b), the spectral changes arose in the same wavenumbers, but its evolution throughout time diverged.

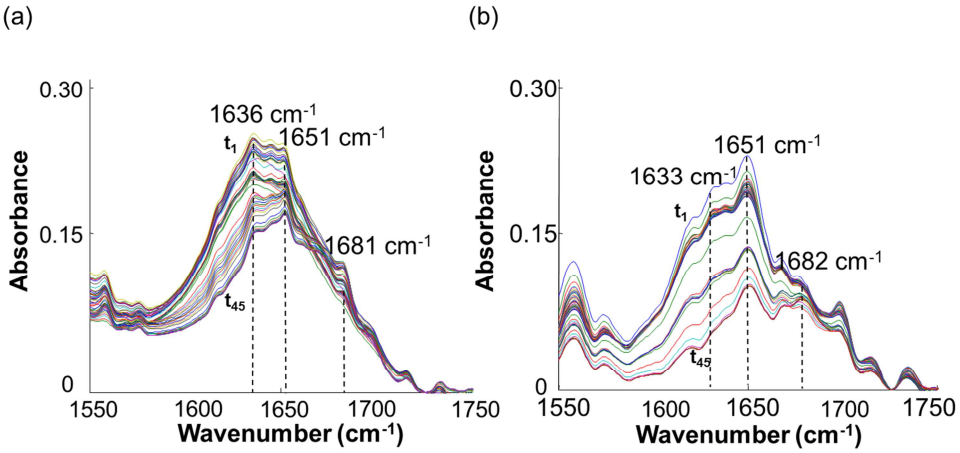


Figure 8. Infrared spectra obtained every minute at 37ºC, during 45 minutes for (a) AZT in serum (control experiment) and (b) insulin: AZT at a stoichiometric ratio in serum.

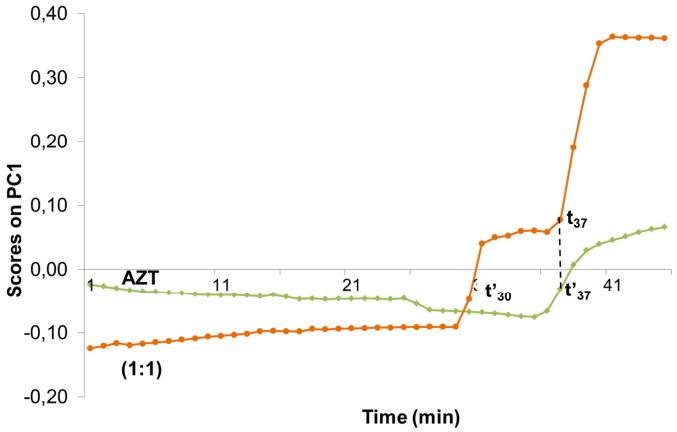


Figure 9. Scores plot on the first PC of (a) AZT in serum (control experiment) and (b) insulin:AZT at a stoichiometric ratio in serum. PC1 accounts for 99 % of the data variability in (a) and 97% in (b).

Afresh, the first scores profile (Fig. 9) obtained when both spectral data matrices were independently analysed by PCA, provided information about the physical changes in the processes of the proteins present in the medium. Whereas a unique transition point at 37 minutes was identified in the control

experiment, related to the aggregates formation, two crossing points at 30 minutes and 37 minutes were seen when insulin was also present. The coincidence between the midpoint at 30 minutes with the observed in the formation of insoluble aggregates in Figure 7b, and the coincidence of the second midpoint and the one in the control experience at 37 minutes led to conclude that the insulin aggregation occurred before the aggregation of serum proteins. Similar conclusions were obtained when the analysis was carried out at higher concentrations of AZT (results not shown). In that case, however, the formation of insulin aggregates appeared in the first 5 minutes and the serum aggregates at 31 minutes, respectively. This fact corroborated that when the AZT concentration in the medium is higher than insulin, the aggregation of insulin is faster.

DISCUSSION

The present work has evidenced the ability of the developed method to monitor insulin aggregation. Significant changes occurred at 38 minutes in solutions of Zn-free insulin at 1.72 mM, which coincides with the aggregates formation. It is most likely that these variations reflect a combination of changes associated with both the dimer dissociation and the oligomerisation of monomers to form aggregates.

In addition, the present work has evidenced the interaction between AZT and insulin. AZT is incorporated in insulin chain forming an “AZT-altered insulin” complex which had an effect in the aggregation. On the one hand, at a concentration of AZT lower than the native insulin considered in the work, the process did not take place during the experimental monitoring time. This fact

was attributed to the decrease in monomer concentration as the amount of the new "AZT-altered insulin" complex increased.

On the other hand, the aggregation was 7 minutes or 35 minutes faster in presence of AZT at a concentration equal or higher than the native insulin, respectively. In addition, two stages in the formation of insulin aggregates were detected. Both stages were related to soluble or insoluble aggregates in function of the AZT concentration level as corroborated by solubility tests.

In the study of the interaction between AZT and insulin in serum medium, the aggregation of serum proteins in presence of AZT was firstly determined. Secondly, this process was analysed in presence of insulin, detecting that the AZT reactivity with insulin was favoured compared with serum proteins.

The commented results can contribute in understanding the relationship between this antiretroviral drug, which is the most widely employed drug in Acquired Immune Deficiency Syndrome (AIDS) treatment, and insulin resistance. Considering that insulin concentration varies along the day and the drugs have an absorption time in human body, the information contained in these results can be of great interest to control the intake of this drug.

ACKNOWLEDGEMENTS. We would like to acknowledge the Universitat Rovira i Virgili, for providing Idoia Martí-Aluja with a doctoral fellowship.

REFERENCES

[1] S. Nakazawa N. Hashii, A. Harazono, N. Kawasak, *Anal. Biochemistry* **420** (2012) 61 – 67.

[2] J. D. Schmit, K. Ghosh, K. Dill, *Biophys. J*. **100** (2011) 450 – 458.

[3] K. Huus, S. Havelund, H. B. Olsen, M. van de Weert, S. Frokjaer, *Biochemistry* **44** (2005) 11171 – 11177.

[4] M. Manno, E. F. Craparo, V. Martorana, D. Bulone, P. L. San Biagio, *Biophys. J.* **90** (2006) 4585 – 1591.

[5] J. Brange, L. Andersen, E. D. Laursen, G. Meyb, E. Rasmussen, *J. Pharm. Sci*. **86** (1997) 517 – 525.

[6] W. Dzwolak, A. Loksztejn, V. Smirnovas, *Biochemistry* **45** (2006) 8143 – 8151.

[7] T. Arakawa, J. S. Philo, D. Ejima, K. Tsumoto, F. Arisaka, *Bioprocess Int.* **4** (2006) 42 – 43.

[8] M. E. M. Cromwell, E. Hilario, F. Jacobson, *AAPS J*. **8** (2006) E572 – E579.

[9] V. Toledo-Rubio, E. Vazquez, G. Platas, J. Domingo-Espín, U. Unzueta, E. Steinkamp, E. García-Fruitós, N. Ferrer-Miralles, A. Villaverde, *J. Biomol. Screen.* **15** (2010) 453 – 457.

[10] R. Jansen, W. Dzwolak, R. Winter, *Biophys J*. **88** (2005) 1344 – 1353.

[11] J. L. Whittingham, D. J. Scott, K. Chance, A. Wilson, J. Finch, J. Brange, G. Dodson, *Mol. Biol.* **318** (2002) 479 – 490.

[12] J. Winter, L. Hauke, R. Rudolph, *Anal. Biochemistry* **310** (2002) 148 – 155.

[13] Ahmad, I.S. Millett, S. Doniach, V. N. Uversky, A. L. Fink, *Biochemistry* **42** (2003) 11404 – 11416.

[14] D. F. Waugh, D. F. Wilhelmson, S. L. Commerford, M. L. Sackler, *J. Am. Chem. Soc*. **75** (1953) 2592 – 2600.

[15] D. F. Waugh, J. Cell. Comp. Physiol. (Suppl 1) **49** (1957) 145 – 164.

[16] C. R. Bloom, R. Heymannn, N. C. Kaarsholm, M. F. Dunn, *Biochemistry* **36** (1997) 12746 – 12758.

[17] L. Shen, J. Shen, X. Luo, F. Cheng, Y. Xu, K. Chen, E. Arnold, J. Ding, H. Jiang, *Biophys. J*. **84** (2003) 3547 – 3563.

[18] M. Aboud, A. Elgalib, R. Kulasegaram, B. Peters, *Int. J. Clin. Pract*. **61** (2007) 463 –472.

[19] Z. He, J. Zhang, X-H. Shi, L-L. Hu, X. Kong X, Y-D. Cai, K-C. Chen, *PLoS ONE* **5** (2010) e9603.

[20] K. C. Chou, *Anal. Biochemistry* **233** (1996) 1 – 14.

[21] L-L. Hu, C. Chen, T. Huang, Y. D. Cai, K-C Chou*, PLoS ONE* **6** (2011) e29491.

[22] L. Chen, W-fM. Zeng, Y-D Cai, K-Y. Feng, K-C. Chou, *PLoS ONE*, **17** (2012) e35254.

[23] K-C. Chou, *Curr. Med. Chem.* **11** (2004) 2105 – 2134.

[24] M. Bouchard, J. Zurdo, E. J. Nettleton, C. M. Dobson,C. V. Robinson, *Protein Sci.* **9** (2000) 1960 – 1967.

[25] W. E. Choi, M. L. Brader, V. Aguilar, N. C. Kaarsholm, M. F. Dunn, *Biochemistry* **32** (1993) 11638 – 11645.

[26] Ahmad, V. N. Uversky, D. Hong, A. L. Fink, *J. Biol. Chem*. **280** (2005) 42669 – 42675.

[27] S. Grudzielanek, R. Jansen, R.Winter, *J. Mol. Biol.* **351** (2005) 879 – 894.

[28] M. Roy, R. W. K. Lee, J. Brange, M. F. Dunn, *J. Biol. Chem*. **265** (1990) 5448 – 5452.

[29] X. Chang, S. Kristensen, J. Led, P. Balschmidt, *Biochemistry* **36** (1997) 9409 – 9422.

[30] C. Ortiz, D. Zhang, Y. Xie, V. Jo Davisson, D. Ben-Amotz, *Anal. Biochemistry* **332** (2004) 245 – 252.

[31] E. J. Nettleton, P. Tito, M. Sunde, M. Bouchard, C. M. Dobson, C.V. Robinson, *Biophys. J*. **79** (2000) 1053-1065.

[32] V. Dzwolak, V. Smirnovas, R. Ansen, R. Winter, *Protein Sci.* **13** (2004) 1927 – 1932.

[33] L. Jørgensen, C. Vermehren, S. Bjerrengaard, S. Froekjaer, *Int. J. Pharm.* **254** (2003) 7 – 10.

[34] Z. Ganim, K. C. Jones, A. Tokmakoff, *Phys. Chem. Chem. Phys.* **12** (2010) 3579 – 3588.

[35] M. del Toro, R. Gargallo, R. Eritja, J. Jaumot, *Anal. Biochemistry* **379** (2008) 8 – 15.

[36] G. Muñoz, A. de Juan*, Anal. Chim. Acta* **595** (2007) 198 –208.

[37] Borges, R. Tauler, A. de Juan, *Anal. Chim. Acta* **544** (2005) 159 – 166.

[38] S. Navea, R. Tauler, A. de Juan, *Anal. Chem.* **78** (2006) 4768 – 4778.

[39] Q-S. Du , Z-Q Jiang, W-Z He, D-P Li, K-C Chou, *J. Biomol. Struct. Dyn*. **23** (2006) 635 – 640.

[40] L. Wang, Y. Wu, F. Meersman*, J. Mol. Struct.* **883** (2008) 79 – 84.

[41] Z-C. Li, X-B. Zhou, Z. Dai, X-Y Zou*, Amino Acids* **37** (2009) 415 –425.

[42] X-L. Liu, J-L. Lu, X-H. Hu*, Protein Peptide Lett*. **18** (2011) 1244 – 1250.

[43] Z-H. Qi, R-Y. Wei, J*. Theor. Biol.* **272** (2011)26 – 34.

[44] MATLAB 6.5, The Mathworks, South Natick, MA, USA (2002).

[45] F. H. Carpenter, *Am. J. Med.* **40** (1966) 750 – 758.

[46] T. N. Yu, C. S. Liu, D. C. O'Shea, *J. Mol. Biol*. **70** (1972) 117 – 132.

[47] S. N. Timasheff, H. Susi, L. Stevens*, J. Biol. Chem*. **242** (1967) 5467 – 5473.

[48] Elliot, E. J. Ambrose, *Nature* **165** (1950) 921 – 922.

[49] S. Krimm, J. Bandekar, *Adv. Prot. Chem*. **38** (1986) 181 – 364.

[50] K. Murayama, M. Tomida, *Biochemistry* **43** (2004) 11526 – 11532.

5.2.2. Results

Quantitative analysis of the effect of zidovudine, efavirenz and ritonavir on insulin aggregation by MCR-ALS of infrared spectra

Idoia Martí-Aluja, Itziar Ruisánchez and M. Soledad Larrechi

Analytica Chimica Acta 760 (2013) 16 – 24

Quantitative analysis of the effect of zidovudine, efavirenz, and ritonavir on insulin aggregation by MCR-ALS of infrared spectra

Abstract. Quantification of the effect of antiretroviral drugs on the insulin aggregation process is an important area of research due to the serious metabolic diseases observed in AIDS patients after ~~a~~ prolonged treatment with these drugs. In this work, multivariate curve resolution alternating least squares (MCR-ALS) was applied infrared monitoring of the insulin aggregation process in the presence of three antiretroviral drugs to quantify their effect. To evidence concentration dependence in this process, mixtures at two different insulin:drug molar ratios were used. The interaction between insulin and each drug was analysed by ^{1}H NMR spectroscopy. In all cases, the aggregation process was monitored during 45 minutes by infrared spectroscopy. The aggregates were further characterised by scanning electron microscopy (SEM). MCR-ALS provided the spectral and concentration profiles of the different insulin-drug conformations that are involved in the process. Their feasible band boundaries were calculated using the MCR-BANDS methodology. The kinetic profiles describe the pathway and the spectral profiles characterise the conformations involved. The retrieved results show that each of the three drugs modifies insulin conformation in a different way, promoting the formation of aggregates. Ritonavir shows the strongest promotion of aggregation, followed by efavirenz and zidovudine. In the studied concentration range, concentration dependence was only observed for zidovudine, with shorter aggregation time obtained as the amount of zidovudine increased. This factor also affected the aggregation pathway.

Keywords – Insulin aggregation; Zidovudine; Efavirenz; Ritonavir; Multivariate curve resolution alternating least squares; FTIR; ^{1}H NMR spectroscopy

INTRODUCTION

Insulin is a small protein consisting of two polypeptides, an A chain (21 amino acids) and a B chain (30 amino acids), linked by two interchain disulphide

bridges [1,2]. Insulin adopts a largely α-helical structure in its native state and rapidly undergoes the aggregation process, losing its native structure [3-5]. As a consequence, the biological activity of insulin is inactivated and becomes harmful to the organism's, leading to disorders including amyloidosis and prion-associated encephalopathies [6,7]. Controlling the aggregation process is an important challenge in the pharmaceutical field, because it affects the stability of insulin when used as a pharmaceutical product [8,9].

There is extensive literature on the insulin aggregation problem, but the effect of antiretroviral drugs on this process is scarcely documented [10]. Prolonged treatment of AIDS patients with antiretroviral drugs, results in metabolic abnormalities such as insulin resistance. Insulin resistance is a disorder in which higher insulin concentrations are required to exert a normal biological response [11,12]. Recently, we have qualitatively reported that insulin remains stable in its native state for less time in the presence of zidovudine, but no information about the quantitative effect of this drug in on the aggregation process has been presented [13].

To reduce the adverse effects of antiretroviral therapy, it is important to quantify the effects of different drugs and to also to understand the underlying mechanism of these effects. The mechanism of insulin aggregation is still under review but is widely accepted to begin with unfolding of insulin, followed by its association (nucleation), to finally obtain the aggregates rich in β-sheet structures [14-16].

There are different types of antiretroviral drugs classified according to the phase of the retrovirus life-cycle that they inhibit. The most common types are: nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), protease inhibitors (PIs) and integrase

inhibitors [17]. The first three groups are the most widely studied, and zidovudine (NRTI), efavirenz (NNRTI) and ritonavir (PI) were selected here.

The aim of the present work was to quantify the effect of antiretroviral drugs in on insulin aggregation, by applying multivariate curve resolution alternating least squares (MCR-ALS) to the infrared spectra recorded during the process. Infrared spectroscopy was chosen to monitor the aggregation process as it is highly sensitive to structural changes of proteins and to the specificity of the amide I region for detecting chemical situations related to the loss of native proteins structure [18-20]. Analyses were carried out at two different insulin:drug molar ratios, to determine the concentration dependence for each drug in this process. MCR-ALS is a well-known methodology that has often been used to understand complicated biological systems, as demonstrated by a series of recent publications [21-27].

The present study was conducted in several stages. First, the interaction between insulin and each drug was proved by ^{1}H nuclear magnetic resonance (NMR) spectroscopy. Comparison of the individual spectrum of insulin with the spectra of the mixtures it formed with each drug showed that new peaks emerged in the amide region in all cases. This indicates that insulin interacts with each drug to form new peptide/amide bonds. The formation of aggregates was characterised by scanning electron microscopy (SEM) in all cases, and a different morphology to that of native insulin aggregates was observed. Second, the effect of each drug was quantitatively assessed by multivariate-curve resolution alternating least squares (MCR-ALS) of the infrared spectra recorded over time. The number of significant species involved in the aggregation process was determined by inspecting the size of the *singular value*, calculated by means of singular value decomposition (SVD). The kinetic concentration profiles retrieved by MCR-ALS indicate that all three drugs

promote aggregation, but that different aggregation pathways exist depending on the drug type and relative concentration. To evaluate the rotational ambiguity of the MCR-ALS solutions recovered in the individual analysis of each experiment, the feasible band boundaries associated with each of the obtained solutions were calculated using the MCR-bands method [28].

METHODOLOGY

Chemical reagents

The drug 3'-azido-3'-deoxythymidine (zidovudine) was purchased in a commercial chemist (GlaxoSmithKline, London). Efavirenz and ritonavir were purchased from Toronto Research Chemicals. CD_3OD was the solvent used for ^1H NMR analysis, purchased from SDS. Zn-free human insulin in a HEPES sodium salt buffer (pH=8.2) at a concentration of 10 mg/ml (1.72mM) was purchased from Sigma-Aldrich and used without further purification. All reagents were used as received. The chemical structure of each drug and a schematic representation of insulin are shown in the supplementary material (Fig. S-1).

GENERAL PROCEDURE

^{1}H NMR spectroscopy

Each antiretroviral drug was mixed at two concentration levels with insulin, to obtain insulin:drug mixtures with a molar ratios of 1:1 and 1:2. The ^1H NMR spectra of native insulin and each drug were also acquired for comparative analysis. Immediately, 500µl of each reaction mixture were transferred to a 5 mm NMR tube, using a double tube system. The external reference tube (o.d. 2 mm, supported by a Teflon adapter) containing the reference substance

(sodium3-trimethylsilyl [2,2,3,3-d4]propionate (TSP) 9.9 mmol/l, $MnSO_4$ 0.47mmol/l in 99.9% D_2O) was placed coaxially in the NMR sample tube (o.d. 5 mm). This double tube system was kept at 4ºC in the sample changer until being analysed. ^1H NMR spectra were recorded at 300 K on a BrukerAvance III 600 spectrometer operating at a proton frequency of 600.20 MHz using a 5 mm CPTCI triple resonance (^1H, ^{13}C, ^{31}P) gradient cryoprobe. One-dimensional ^1H pulse experiments were carried out using the nuclear Overhauser effect spectroscopy (NOESY)-presaturation sequence (RD-90º–t1–90º–tm–90º ACQ) to suppress the residual water peak. t_1 time was set at 4 ms and tm (mixing time) at 100 ms. The 90° pulse length was calibrated for each sample and varied from 7.18 µs to 7.31 µs. The spectral width was 20 ppm, and a total of 64 transients were collected into 64 k data points for each ^1H spectrum.

FTIR-ATR

The same experimental procedure was used to prepare the mixtures (insulin-drug) for infrared monitoring. In this case, to take the measurements a drop of the mixture was placed on a small diamond crystal in the spectrophotometer ATR cell (FTIR 680 Plus JASCO and a RS232 Control), which was continuously purged with N_2. The FTIR spectra were recorded *in situ* every minute, for 45 minutes, in the spectral range 1550-1750 cm^{-1}. The CO_2 contribution was removed with the control software Spectra Manager before the spectra were exported to Matlab [29]. All experiments were carried out at 37ºC, to mimic human body conditions. A solution of human insulin was also monitored using the same experimental conditions to provide comparative analysis. In addition, the infrared spectrum of each drug dissolved in the insulin buffer (HEPES) was acquired. The infrared spectrum of pure zidovudine was recorded during 45 minutes to evaluate the noise.

Multivariate curve resolution-alternating least squares (MCR-ALS)

The 45 infrared spectra from each experiment were arranged as rows in **D** matrices of dimensions 45 x 209. First, Savitzky-Golay smoothing was applied to all data sets to suppress instrumental noise. Assuming a bilinear model, the matrix **D** can be decomposed into the product of two matrices according to equation 1:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^{T} + \mathbf{E} \qquad (1)$$

where **C** contains the concentration profiles of the different species and $\mathbf{S}^{T}$ their related pure spectra. **E** is the experimental error unexplained by the resolution model.

The main goal of MCR-ALS is to determine **C** and $\mathbf{S}^{T}$ matrices solely from the information of the experimental data matrix **D** [30]. MCR-ALS works following several stages. First, the number of species involved in the aggregation process must be determined. This determination is performed by Singular Value Decomposition (SVD) [31], based on the assumption that the *singular values* related to chemical species are larger than those linked to experimental errors and noise. In this study, we consider the second *singular value* calculated with the data matrix of the reagent (AZT dissolved in HEPES) as the noise threshold value. Second, the initial estimates of $\mathbf{S}^{T}$ or **C** are built by evolving factor analysis (EFA) [32].These initial estimates are then iteratively optimised by a constrained alternating least squares (ALS) regression procedure. New estimates of the spectral matrix ($\mathbf{S}^{T}$) and the concentration matrix (**C**) are obtained at each iteration.

The quality of MCR-ALS is evaluated using the lack of fit (LOF) and explained variance (R^2) parameters, which identify the dissimilarity between the experimental data matrix (**D**) and the data reproduced by MCR-ALS:

$$LOF\ (\%) = \ 100 \cdot \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \qquad (2)$$

$$R^2(\%) = \left(1 - \frac{\sum_{i,j} d_{ij}^2 - \sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}\right) \cdot 100 \quad (3)$$

where d_{ij} is an element of the input data matrix **D** and e_{ij} is the related residual obtained from the difference between the experimental data matrix (**D**) and the MCR-ALS reproduction.

Two different LOF values are calculated, according to the defined input data matrix **D** (taken either as the raw experimental data matrix or the PCA reproduced data matrix, using the same number of components as the MCR-ALS model). The closer the values of LOF and R^2 to zero and one, respectively, the better the fit of the experimental data.

The similarity coefficient (r), which represents the correlation between the recovered profiles and the real profiles, was used to evaluate the quality of the spectral profiles:

$$r = cos\ \gamma = \frac{s_i^T \hat{s}_i}{\|s_i\| \cdot \|\hat{s}_i\|} \qquad (4)$$

where γ is the angle of the vectors associated with the profile recovered by MCR-ALS ($\hat{s}_i$) and the real profile (s_i), for each species *i*.

Equation (1) has multiple solutions. This problem is called factor analysis ambiguity [33] and is mathematically described by the following equation:

$$\mathbf{D} = \mathbf{C}_{old}\,\mathbf{S}_{old}^{T} = (\mathbf{C}_{old}\,\mathbf{T}^{-1})\,(\mathbf{T}^{-1}\,\mathbf{S}_{old}^{T}) = \mathbf{C}_{new}\,\mathbf{S}_{new} \qquad (5)$$

For a particular species profile, the set of feasible solutions under constraints defines a range or band of feasible solutions, and this band may be delimited by maximum and minimum band boundaries In this study, the boundaries were defined using the MCR-BANDS method [28], which obtains them from a constrained non-linear optimisation of an objective function $f_n(\mathbf{T})$ that depends on the rotation matrix **T** (Eq. 6). This function is scaled between 0 and 1 and provides an estimation of the contribution of each component profile signal to the overall signal contribution.

$$f_n(\mathbf{T}) = \frac{\left\|\mathrm{c_n s_n^T}\right\|}{\|\mathbf{C}\mathbf{S}^{\mathbf{T}}\|} \qquad (6)$$

Scanning electron microscopy (SEM)

SEM micrographs were taken in each experiment using a JEOL JSM-6400 SEM microscope, operating at 15 KV. Samples were coated with gold before examination (cathode dispersion).

RESULTS

^{1}H NMR spectral changes associated with the interaction between drugs and insulin

Figure 1 shows the ^{1}H NMR spectra of native insulin and the mixtures of insulin with the different drugs at a 1:1 molar ratio. The ^{1}H NMR spectra of the pure drugs dissolved in $CDCl_3$ are included in the supplementary material (Fig. S-2). All drugs showed peaks in the range 0–7 ppm of the NMR spectra, as was

expected according to its structure. Nevertheless, ritonavir showed a band at around 8.8 ppm, which can be attributed to the NH groups present in its structure (see Fig. S-1). The spectra of each insulin-drug mixture (Figs. 1B, 1B and 1C) show peaks not found in the native insulin spectrum (Fig. 1A), indicating that the insulin structure has changed. Thus, a new structural form of insulin is present in the mixtures, which will be referred to as “insulin-modified” structure in this paper. For insulin-zidovudine (Fig. 1B), a new peak appeared at 9.1 ppm, which was not present in the insulin spectrum (Fig. 1A) or the zidovudine spectrum (Fig. S-2). This peak is characteristic of NH amide protons and can be attributed to the presence of amide bonds, which may be formed from the opening of the uracil group in zidovudine by attack of final amino groups of insulin (Fig. S-1) [33-37]. The presence of the 7.6 ppm band, which corresponds to the NH of the uracil group (seen in the zidovudine spectrum, Fig. S-2), suggests that the uracil is only partially opened [34].

Comparison of the insulin-efavirenz (Fig. 1C) and insulin-ritonavir (Fig. 1D) spectra with the respective drug spectra (Fig. S-2) reveals a new peak at 12.4 ppm, which is characteristic of hydrogen bonding [35, 36]. Therefore, this peak may be related to the formation of hydrogen bonds between insulin and each drug. This interaction indicates that the conformation of insulin has changed in the presence of efavirenz and ritonavir, leaving the non-polar residues of insulin inside. This means that the hydrophilic regions of insulin are freely available to form new hydrogen bonds.

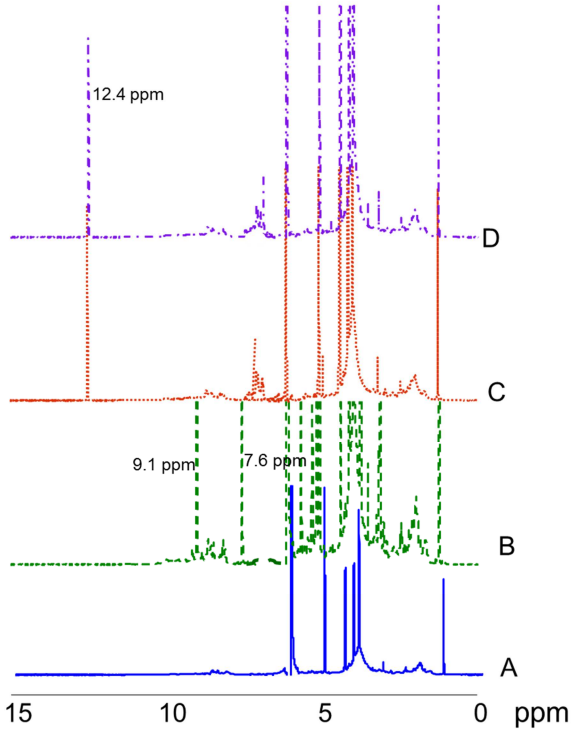


Figure 1. ^{1}H NMR spectrum of native insulin (A, solid line), insulin-zidovudine (B, dashed line), insulin-efavirenz (C, dotted line) and insulin-ritonavir (D, dotted-dashed line).

Clearer evidence of the interaction between drugs and insulin can be obtained from more detailed analysis of the spectra. The ^{1}H NMR spectrum of an individual protein can be divided into different regions; the ones that give the most information are the NH backbone region (8-12 ppm) and the H_{α} region (3-5 ppm) [36, 37]. The individual residues of a protein structure are packed into chemical environments, which are shifted in the event of a structural perturbation. The chemical shifts of the ^{1}Hα of the peptide backbone and the final amino protons (Gly, Phe, Asn and Thr) of insulin are commonly used to examine the modification of native insulin structure [38,39]. These regions are enlarged in Figure 2a and Figure 2b, respectively Information associated with

the ^{1}H NMR of the dissolved drug is shown in Figure S-3. Figure 2a shows the spectral region of ^{1}Hα for insulin (A), insulin-zidovudine (B), insulin-efavirenz (C) and insulin-ritonavir (D). A slight shift of the peaks is detected in the presence of each drug in the medium, the ^{1}Hα peak of Gly showing the greatest displacement. Although some peaks in the ritonavir and zidovudine spectra match those of amino acids (e.g., 3.89 ppm or 4.35 ppm, Fig. S-3), their shapes are different to those observed in the insulin and insulin-drug spectra. Thus, the ^{1}Hα chemical shift observed when insulin is mixed with the drugs (Fig. 2A) may be attributable to the change in insulin structure as a result of the drug interaction. In addition, the insulin-zidovudine spectrum shows a new, broad peak at 3.75 ppm, by possibly due to the new methine proton (-CH-) generated in the uracil opening described above. This suggests that the native structure of insulin has changed; the Gly environment shows the greatest disruption, which is consistent with the chemical shifts of the final amino protons (Fig. 2b). Signals between 8 and 9 ppm, which correspond to the NH and NH_2 of the final amino acids [38,39], are also shifted in the insulin-drug spectra when compared to the native insulin spectrum. Thus, the chemical environment surrounding the final amino acids is disturbed by the presence of each drug. From this analysis, it was assumed that each drug joins the insulin chain, modifying the native structure of insulin. The same behaviour was observed in all experiments at 1:2 molar ratios (results not shown), where the displacement was more pronounced than at 1:1 ratios.

FTIR and SEM analysis

As stated above, the formation of aggregates in the presence of the three drugs was followed by FTIR-ATR for 45 minutes. Figure 3 shows the initial (solid line) and final (dotted line) spectra for insulin and for each mixture. The spectra

have shifted up the y-axis to clarify the discussion. At this point in the study, it was assumed that the initial spectra shown in Figures 3B, 3C and 3D are representative of an "insulin-modified" sample. As the main objective of our project is to extract quantitative information about the effect of each drug on insulin aggregation, the characteristic bands have been marked in the spectra: the absorption band at 1634 cm^{-1}, characteristic of the α-helix structure in native insulin; and the absorption bands at 1617 cm^{-1}, 1625 cm^{-1}, 1652 cm^{-1} and 1683 cm^{-1}, which are linked to intermolecular β-sheets and β-turn structures of partial unfolding of insulin [40-45].

The initial spectrum of each sample ("insulin-modified", Figs. 3B, 3C and 3D) did not vary significantly from the corresponding spectrum of each drug when dissolved in an HEPES solution (see, for example, Fig. S-4 which shows the spectrum of zidovudine). This is unsurprising, since in the same region it is also possible to observe the stretching vibration of C=O functional groups present in all drugs (Fig. S-1). Although the carbonyl groups of each drug have different environment, the infrared spectra of each drug diluted in HEPES were very similar to the one shown in Figure S-4. This may be attributable to the formation of hydrogen bonds between the hydroxyl groups of the HEPES solvent and the carbonyl groups of the drugs, resulting in a broad absorption band as shown in Figure S-4. The slight differences between the initial spectrum shown in Figure 3 and the drug spectrum were more pronounced in the zidovudine experiment (Fig. 3B). This is explained by taking into account that the spectral region can be expected to show the absorption bands of the C=O functional groups corresponding to the new amide groups formed in the "insulin-modified" structure by the interaction between insulin and zidovudine.

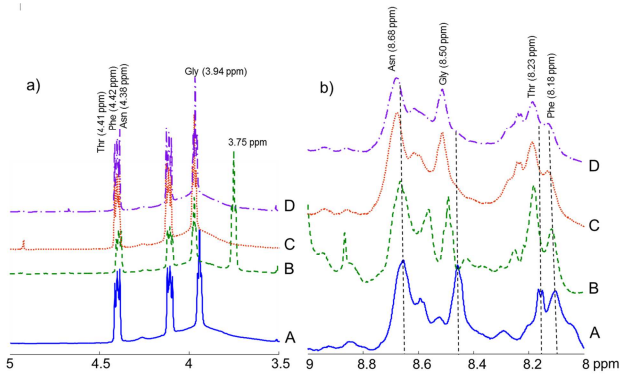


Figure 2. Enlarged ^{1}H NMR spectrum of native insulin (A, solid line), insulin-zidovudine (B, dashed line), insulin-efavirenz (C, dotted line) and insulin-ritonavir (D, dotted-dashed line) in the a) $^{1}H_{\alpha}$ region and b) final amino protons region.

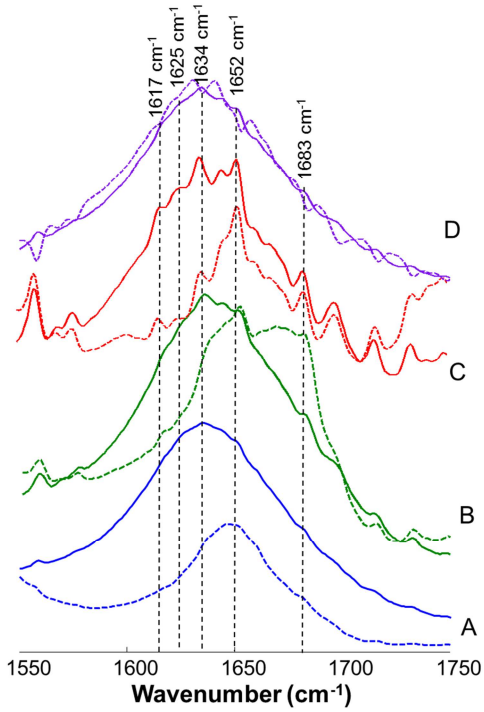


Figure 3. First (solid line) and last (dotted line) infrared spectrum for native insulin (A, solid line), insulin-zidovudine (B, dashed line), insulin-efavirenz (C, dotted line) and insulin-ritonavir (D, dotted-dashed line).

A comparative analysis of the initial spectra recorded for each experiment (Figs. 3B, 3C and 3D) provides preliminary information about the way in which each drug alters the native structure of insulin. In the efavirenz experiment (Fig. 3C), the characteristic β-structure absorption peaks at 1617 cm^{-1}, 1625 cm^{-1} and 1683 cm^{-1} are better defined than in the zidovudine experiment (Fig. 3A), suggesting that greater modification of the native insulin structure occurs. This is consistent with the explanation given in the previous section concerning the formation of hydrogen bonds between efavirenz and insulin. Similarly, the broad band observed in the ritonavir experiment (Fig. 3D) is consistent with the favoured formation of hydrogen bonds due to the chemical structure of ritonavir [40-45]. Comparison of the initial and final spectra of all experiments (Fig. 3) clearly shows that the spectral pattern of aggregation differs according to the drug administered. This effect is clearly demonstrated by the chemometrics analysis of the infrared data matrices, which will be discussed later.

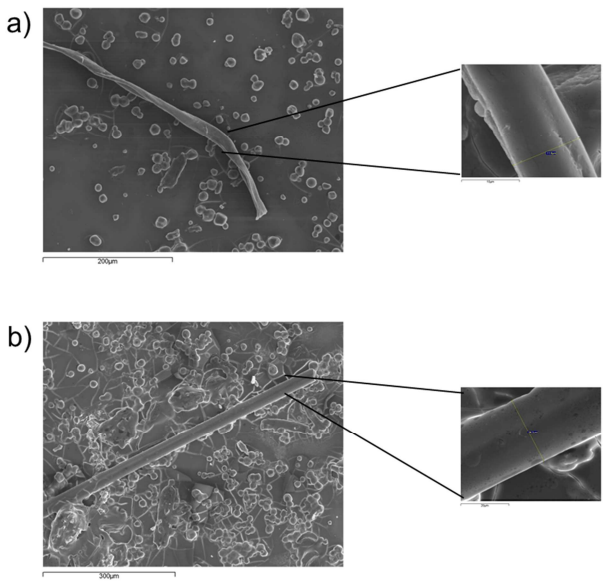


Figure 4. SEM images of insulin:zidovudine at a) (1:1) molar ratio and b) (1:2) molar ratio.

The overall profile for insulin:zidovudine at a 1:2 molar ratio was different to the one shown in Figure 3B, despite the fact that the same spectral features are observed. The amide I band was broader, which suggests that a greater presence of unfolded parts in the "insulin-modified" structure from the beginning of the experiment. The other experiments at this molar ratio presented an insoluble material from the beginning, which indicates the presence of aggregates. At this point, the experiments at 1:2 molar ratios were discarded for further analysis.-Visal evidence of the effect of each drug on the aggregation process was obtained by examining the SEM images of the "insulin-modified", captured after 45 minutes in each experiment. The images show that the fibrils formed were different, as can be seen in Figure 4 for insulin:zidovudine at both molar ratios (1:1 and 1:2). Distinct morphological stages are observed in these images. In the first experiment, small and ordered clusters of aggregates / fibres, with an approximate diameter of 11.9 µm, were formed (Fig. 4a, 1:1); these fibres did not have the regular fibrillar appearance found in aggregates formed by native insulin. In the second experiment, amorphous aggregated clusters with an approximate diameter of 20.2 µm were found to be distributed along the surface (Fig. 4b). Therefore, it can be postulated that the structure of the modified insulin (insulin-drug) is different in each mixture. Consequently, the quantitative analysis of the effect of each drug on the aggregation pathway spectra was performed independently by MCR-ALS analysis of the infrared data.

MCR-ALS and MCR-bands analysis

Table 1 shows the first five *singular values* associated with the SVD of all spectral data sets. The explained variance (%) of the original data is given in

brackets. In all cases, almost 100% of the variance is explained by these five factors.

Table 1. Rank analysis of each experiment. The first five *singular values* are shown, and the explained variance (%) of the original data is indicated in parenthesis.

	Singular values					
Factor number	**Insulin: zidovudine**		**Insulin: efavirenz**		**Insulin: ritonavir**	
	(1:1)	(1:2)	(1:1)	(1:2)	(1:1)	(1:2)
1	**7,60 (90,24%)**	**6,07 (93,60%)**	**3,66 (99,54%)**	**2,40 (99,62%)**	**5,00 (98,10%)**	**3,20 (99,20%)**
2	**0,93 (94,18%)**	**0,35 (97,49%)**	**1,10 (99,73%)**	**0,20 (99,96%)**	**0,14 (99,63%)**	**0,10 (99,73%)**
3	**0,11 (98,22%)**	**0,11 (99,73%)**	0,05 (99,96%)	0,04 (99,98%)	0,01 (99,74%)	0,01 (99,98%)
4	0,04 (99,90%).	0,02 (99,84%)	0,03 (99,98%)	0,02 (99,99%)	0,01 (99,82%)	0,01 (99,99%)
5	0,03 (99,99%)	0,01 (99,94%)	0,01 (99,99%)	0,01 (99,99%)	0,00 (99,98%)	0,00 (99,99%)

The number of retained factors before MCR-ALS analysis was estimated by examining the magnitude of *singular values*. In this study, 0.0435 was taken as the threshold value for being representative of noise in the experimental conditions employed. This value corresponds to the second singular value from the SVD analysis of the data matrix containing zidovudine spectrum recorded for 45 minutes. Based on this criterion, three factors were retained for the 1:1 and 1:2 insulin:zidovudine experiments, which contained 98.2% and 99.7% of the variance, respectively. By contrast, two factors were retained for the other experiments, which accounted for over 99.6% of the total variance. The concentration profiles obtained by EFA were used as initial estimates for the MCR-ALS analysis. The constraints imposed during the optimisation step were non-negativity and normalisation for the spectral profiles and non-negativity for the concentration profiles. Local selectivity constraints were imposed for the first chemical species during the first 15 minutes for insulin:zidovudine at

1:1. In all cases, the similarity coefficient (Eq. 4) between the initial spectrum recorded for each experiment – representative of the initial "insulin-modified" species – and the spectrum recovered by MCR-ALS for the species was calculated. A value above 0.98 was obtained in all cases. Lack of fit (Eq. 2) and residual variance (Eq. 3) were also calculated, with values below 1.8% and equal to or above 99%, respectively, obtained in all cases.

Figures 5–7, below, show some of the results obtained. The concentration and spectral profiles for insulin:zidovudine at 1:1 molar ratio (Figs. 5a and 5b) fit an aggregation scenario involving the presence of an intermediate (species 2) between the initial species 1 and the final species 3.

As can be expected, assuming that the insulin structure has been modified in the samples, the recovered spectrum for each species (Fig. 5b) contains the spectral features of the β-structures. However, the intensity values of the bands around 1683 cm^{-1} (representative of β-turn structures) are higher for species 2 than species 1, but lower than for species 3. This can be explained by the fact that the environment of the sample became more rigid over time, reducing the stretching vibration of the active functional groups. Global consideration of the results leads us to postulate that the aggregation process develops as shown below in Scheme 1:

$$\langle \text{insulin} - \text{zidovudine} \rangle_{modified\ insulin} \rightarrow \text{nucleus intermediate} \rightarrow \text{aggregates}$$

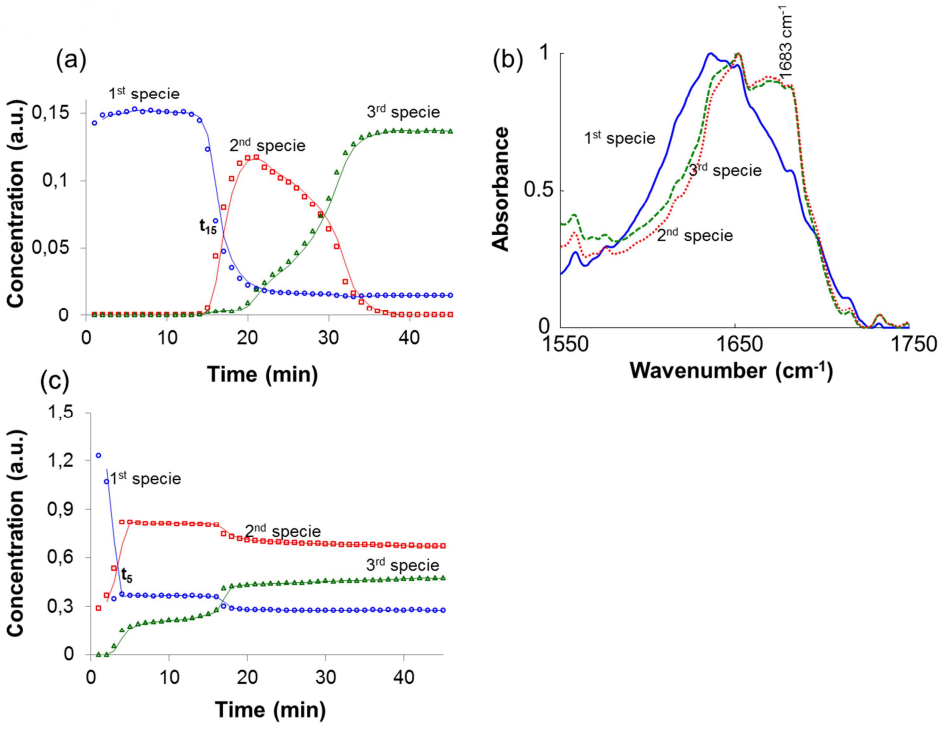


Figure 5. (a) Concentration profiles and (b) spectral profiles for 1:1 insulin:zidovudine. (c) Concentration profiles for 1:2 insulin:zidovudine (o solid line: first specie, dashed line: second specie and Δ dotted line: third specie).

The evolution of the concentration profiles (Fig. 5a) suggests that nucleus formation occurred during the first 15 minutes, in which the initial "insulin-modified" sample slowly lost its native structure. Once a large number of nuclei had been formed (species 2, at 20 minutes), formation of aggregates (species 3) was favoured. The nucleus intermediates are mainly composed by a mixture of α-helix, β-turn and random structures, whereas the aggregates present only a β-sheet conformation.

The recovered spectra in the MCR-ALS analysis of the data matrix (insulin: AZT at molar ratio 1:2) were similar to those shown in Figure 5a. Thus, the spectral

considerations are similar to those for the 1:1 ratio experiment. However, the evolution of the concentration profiles shown in Figure 5c suggests that the sample can be considered as a mixture of two different species. This statement is chemically sound, since the amount of uracil groups is higher in the 1:2 experiment, meaning that uracil opening could come not only from the terminal amino groups but also from adjacent groups. Both options lead to the formation of insulin fractions with a structure that exhibits varying degrees of modification, which for the purposes of this study are referred to as "partially unfolded" or "totally unfolded". Given these considerations, it is possible to postulate Scheme 2, which represents aggregation of insulin in the presence of zidovudine at a molar ratio of 1:2:

$$\langle \text{insulin} - \text{zidovudine} \rangle \begin{matrix} \nearrow \\ \searrow \end{matrix} \begin{matrix} \text{partially unfolded} \\ \text{totally unfolded} \end{matrix} \rightarrow \text{aggregates}$$

Schemes 1 and 2 are consistent with the captured SEM images (Fig. 4). In the 1:2 experiment, different species are involved in the process, leading to thicker and grouped aggregates, as shown in Figure 4b. By contrast, in the 1:1 image, small and ordered clusters of aggregates with smaller diameters are observed (Fig. 4a).

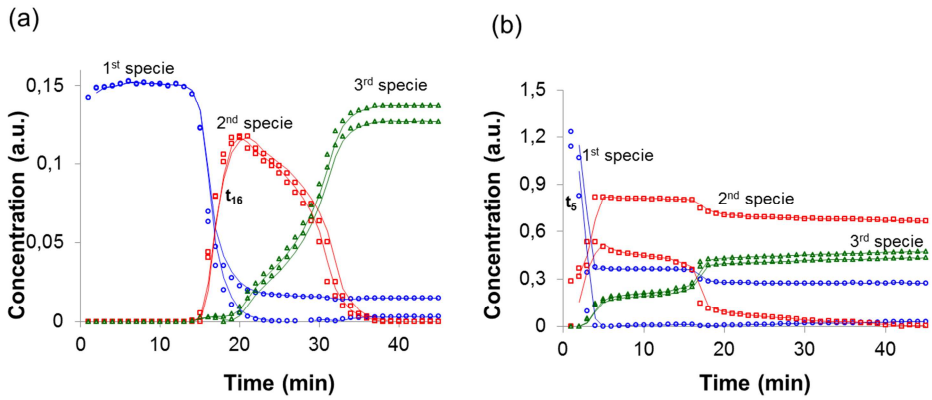


Figure 6. Concentration profiles corresponding to the maximum and minimum component

relative contribution optimization function f_n^{min} and f_n^{max} for: (a) 1:1 insulin:zidovudine and (b) 1:2 insulin:zidovudine (o solid line: first specie, dashed line: second specie and Δ dotted line: third specie).

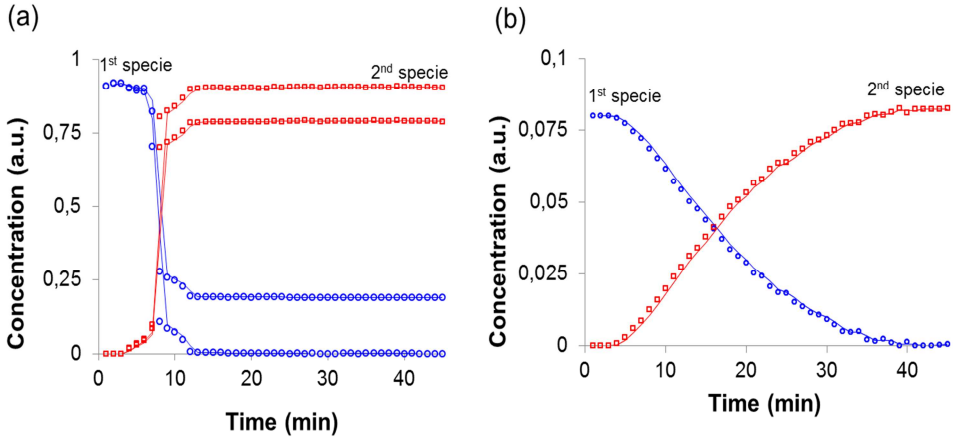


Figure 7. Concentration profiles corresponding to the maximum and minimum component relative contribution optimization function (f_n^{min} and f_n^{max}) at 1:1 molar ratio for a) insulin:efavirenz and b) insulin:ritonavir (o solid line: first specie and, dashed line: second specie).

Nevertheless, it is known that MCR-ALS solutions of a single matrix suffer from rotational ambiguity. To evaluate the evolution of the aggregation process more accurately, the feasible band boundaries associated with each of the solutions were calculated using the MCR-BANDS method. As the spectral profiles of the different compounds do not differ significantly, only the f_n^{max} and f_n^{min} values (Eq. 6) of the concentration profiles are shown in Figure 6. In this case, the ambiguity of the solutions is relatively low for the first experiment (insulin:zidovudine at 1:1, Fig. 6a) and can be seen after 16 minutes. Therefore, a more realistic approximation of the aggregation pathway would be the one shown in Scheme 3:

$$\langle \text{insulin} - \text{zidovudine} \rangle_{modified\ insulin} \rightleftarrows \text{nucleus intermediate} \leftrightarrows \text{aggregates}$$

The ambiguity was higher for the 1:2 experiment (Fig. 6b). In this case, two different estimations can be extracted according to the band limit observed: on the one hand, insulin has practically lost its partially unfolded form (species 1) at 5 minutes; on the other, an important amount of partially unfolded forms remains at the end of the experiment (see specie 1). Consequently, a more realistic approximation of the aggregation process would be as shown in Scheme 4:

$$\langle \text{insulin} - \text{zidovudine} \rangle \begin{matrix} \nearrow \\ \searrow \end{matrix} \begin{matrix} \text{partially unfolded} \\ \text{totally unfolded} \end{matrix} \quad \rightleftarrows \quad \text{aggregates}$$

Figures 7a and 7b show the concentration profiles and feasible bands recovered from the efavirenz and ritonavir experiments at a molar ratio of 1:1. The spectral profiles are similar to those shown in Figure 3. As explained in the previous section, spectroscopic analysis suggested a higher presence of β-structures from almost the beginning of the experiment (initial spectrum in Figs. 3C and 3D). Thus, it can be assumed that, in both cases, the structure of the sample is strongly modified. This suggests the presence of an aggregation nucleus from the beginning of the process. The evolution of the concentration profiles (Figs. 7A and 7B) suggests that the aggregation process follows the pathways shown below in Scheme 5 and Scheme 6:

$$\langle \text{insulin} - \text{efavirenz} \rangle_{nucleus} \rightleftarrows \text{aggregates}$$

$$\langle \text{insulin} - \text{ritonavir} \rangle_{nucleus} \rightarrow \text{aggregates}$$

The feasibility bands around the solutions of Figure 7a indicate that the aggregation state of insulin-efavirenz is lower than insulin-ritonavir sample at the end of the process (45 minutes). This statement is in agreement with the

expected results, considering the higher capacity of ritonavir to form hydrogen bonds with insulin (see infrared section).

CONCLUSIONS

There is clear evidence that all three of the antiretroviral drugs analysed in this paper (zidovudine, efavirenz and ritonavir) interact with insulin. However, the methodology presented has demonstrated that each drug modifies the native conformation of insulin in a different way.

Analysis of infrared data using the MCR-ALS method has been shown to be a powerful technique for obtaining quantitative evidence of the effect of antiretroviral drugs on the natural process of insulin aggregation and for extracting additional information about the pathway along which the process occurs.

The incorporation of the feasibility bands calculated by MCR-BANDS enabled more accurate conclusions to be drawn from the results.

The three drugs analysed promote aggregation. The greatest acceleration is caused by ritonavir, followed by efavirenz and zidovudine.

In the experimental conditions considered, concentration dependence was only observed for zidovudine. Furthermore, two effects were observed: a) the formation of aggregates occurred earlier at the highest level of zidovudine, which may help to understand why insulin disorders in HIV patients are detected after ~~a~~ prolonged treatment with antiretroviral drugs; b) two different aggregation pathways were postulated, depending on the drug concentration level. Two consecutive stages were involved in the process at the lowest molar

ratio, while two parallel pathways were postulated for the highest molar ratio. This information could be used as a guideline for in-depth mechanistic studies.

ACKNOWLEDGEMENTS. We would like to acknowledge the Universitat Rovira i Virgili, for providing Idoia Martí-Aluja with a doctoral fellowship.

REFERENCES

[1] D-P. Hong, A. Ahmad, A. L. Fink, *Biochemistry* **45** (2006) 9342 – 9353.

[2] U. Derewenda, Z. Derewenda, G. G. Dodson, R. E. Hubbard, F. Korber, *Brit. Med. Bull.* **45** (1989) 4 – 18.

[3] J. Brange, L. Andersen, E. D. Laursen, G. Meyb, E. Rasmussen, *J. Pharm. Sci.* **86** (1997) 517 – 525.

[4] Q. Hua, M. A. Weiss*, J. Biol. Chem*. **279** (2004) 21449 – 21460.

[5] A. Ahmad, V. N. Uversky, D. Hong, A. L. Fink, *J. Biol. Chem.* **280** (2005) 42669 –42675.

[6] M. Dobson, *Trends Biochemistry Sci*. **24** (1999) 329 – 332.

[7] M. Bucciantini, E. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dovson, M. Stefani, *Nature* **416** (2002) 507 – 511.

[8] K. Giger, R. P. Vanam, E. Seyrek, P. L. Dubin, *Biomacromolecules* **9** (2008) 2338 –2344.

[9] M. R. DeFilippis, R. E. Chance, B. H. Frank, *Crit. Rev. Ther. Drug Carrier Syst*. **18** (2001) 201 – 264.

[10] J. Ghosn, M.L. Chaix, *AIDS* **24** (2010) 309 – 311.

[11] M. Aboud, A. Elgalib, R. Kulasegaram, B. Peters, *Int. J. Clin. Pract.* **61** (2007) 463 –472.

[12] S. B. Biddinger, C. R. Kahn, *Annu. Rev. Physiol*. **68** (2006) 123 – 158.

[13] I. Martí-Aluja, J. Ferré, I. Ruisánchez, M. S. Larrechi, *Chemom. Intell. Lab. Syst.* (2012) doi: 10.016/j.chemolab.2012.06-007.

[14] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky, A. L. Fink, *Biochemistry* **40** (2001) 6036 – 6046.

[15] L. Skopra, S. Becker, M. Zweckstetter*, J. Am. Chem. Soc* **132** (2010) 9223 – 9225.

[16] L Nielsen, S. Frokjaer, J. Brange, V. N. Uversky, A. L. Fink, *Biochemistry* **40** (2001) 8397 – 8409.

[17] A. Checa, R. Oliver, S. Hernández-Cassou, J. Saurina, *Anal. Chim. Acta* **647** (2009) 1 – 13.

[18] V.Dzwolak, V. Smirnovas, R. Ansen, R. Winter*, Protein Sci.* **13** (2004) 1927 – 1932.

[19] L. Jørgensen, C. Vermehren, S. Bjerrengaard, S. Froekjaer, *Int. J. Pharm*. **254** (2003) 7 – 10.

[20] Z. Ganim, K. C.Jones, A. Tokmakoff, *Phys. Chem. Chem. Phys.* **12** (2010) 3579 –3588.

[21] S. Ruiz-Castelar, A. Checa, R. Gargallo, J. Jaumot, *Anal. Chim. Acta* **722** (2012) 34 –42.

[22] J. Saurina, S. Hernández-Cassou, R. Tauler, A. Izquierdo-Ridorsa, *Anal. Chem*. **71** (1999) 126 – 134.

[23] J. M. Amigo, A. de Juan, J. Coello, S. Maspoch*, Anal. Chim. Acta* **567** (2006) 236 –244.

[24] G. Muñoz, A. de Juan, *Anal. Chim. Acta* **595** (2007) 198 – 208.

[25] A. Borges, R. Tauler, A. de Juan, *Anal. Chim. Acta* **544** (2005) 159 – 166.

[26] S. Navea, R. Tauler, A. de Juan, *Anal. Chem*. **78** (2006) 4768 – 4778.

[27] A. Checa, R. Oliver, J. Saurina, S. Hernández-Cassou, *Anal. Chim. Acta* **572** (2006) 155 – 164.

[28] J. Jaumot, R. Tauler, *Chemom. Intell. Lab. Syst*. **103** (2010) 96 – 107.

[29] The Mathworks, MATLAB Version 7.0, Natick, MA, 2004.

[30] R. Tauler, *Chemom*. *Intell*. *Lab*. *Syst*. **30** (1995) 133 – 146.

[31] D. L. Massart, B. Vandeginste, L. Buydens, S. de Jong, P. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A*, Elsevier, Amsterdam, 1997.

[32] M. Maeder, *Anal. Chem*. **59** (1987) 527 – 530.

[33] M. Garrido, M. S. Larrechi, F. X. Rius, R. Tauler, *J. Chemom.* **76** (2005) 111 – 120.

[34] E. Pretsch, P. Bühlmann, M. Badertscher, Springer (Eds.), *Structure Determination of Organic Compounds: Tables of Spectral Data*, Springer, Berlin, 2009.

[35] L. Pauling, R. B. Corey, H. R. Branson, *PNAS*.**37** (1951) 205 – 211.

[36] H. J. Dyson, P. E. Wright, *Annu. Rev. Phys. Chem*. **47** (1996) 369 – 395.

[37] C. R. Sanders, F. Sonnichsen, *Magn. Reson. Chem*. **44** (2006) S24 – S40.

[38] A. D. Kline, R. M. Justice Jr., *Biochemistry* **29** (1990) 2906 – 2913.

[39] Y. Yoshimura, K. Sakurai, Y-H. Lee, T. Ikegami, E. Chatani, H. Naiki, Y. Goto, *Protein Sci.* **19** (2010) 2347 – 2355.

[40] M. Bouchard, J. Zurdo, E. J. Nettleton, C. M. Dobson , C. V. Robinson, *Protein Sci.* **9** (2000) 1960 – 1967.

[41] S. N. Timasheff, H. Susi, L. Stevens, *J. Biol. Chem.* **242** (1967) 5467 – 5473.

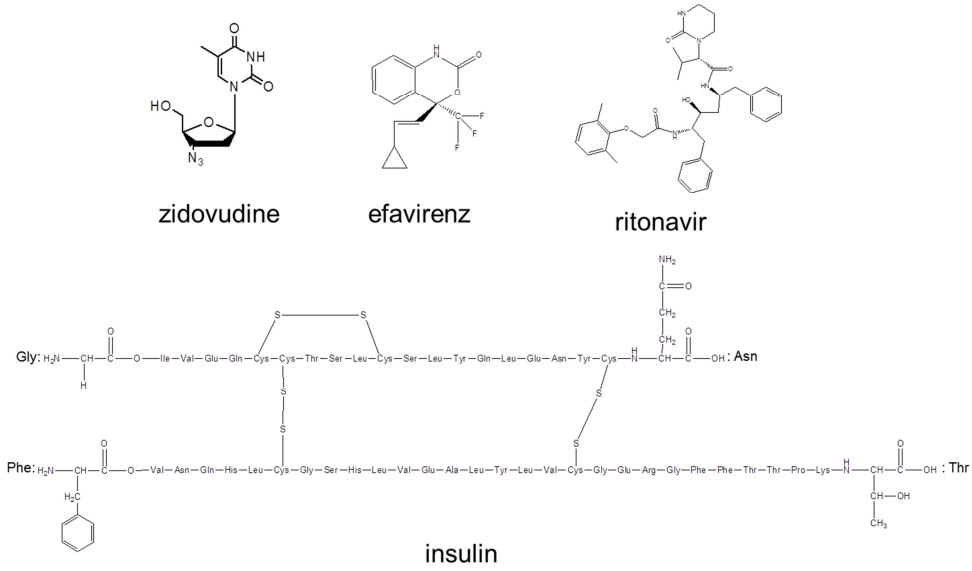
[42] A. Elliot, E. J. Ambrose, *Nature* **165** (1950) 921 – 922.

[43] S. Krimm, J. Bandekar, *Adv. Prot. Chem*. **38** (1986) 181 – 364.

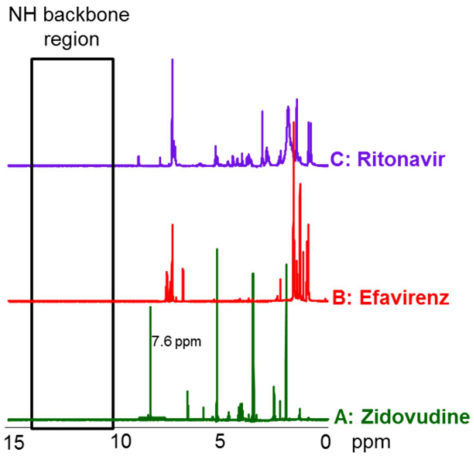
[44] K. Murayama, M. Tomida, *Biochemistry* **43** (2004) 11526 – 11532.

[45] E. S. Manas, Z. Getahun, W. W. Wright, W. F. DeGrado, J. M. Vanderkooi, *J. Am. Chem. Soc.* **122** (2000) 9883 – 9890.

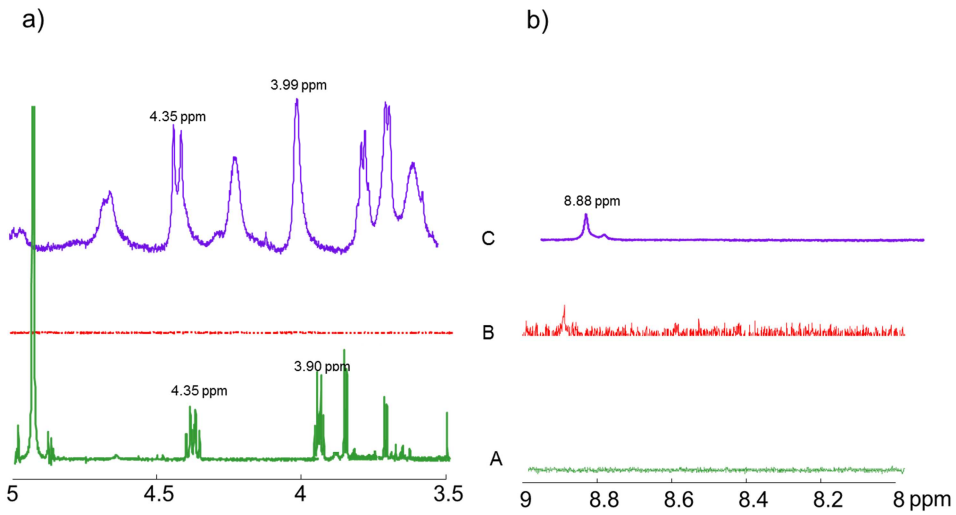
Supplementary material



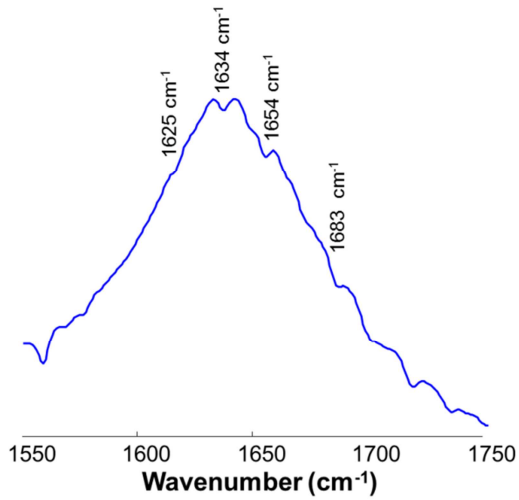
Supplementary 1. Chemical structure of insulin, zidovudine, efavirenz and ritonavir.



Supplementary 2. ^{1}H NMR spectrum of zidovudine (A, green line), efavirenz (B, red line) and ritonavir (C, purple line).



Supplementary 3. Enlarged ^{1}H NMR spectrum of zidovudine (A), efavirenz (B) and ritonavir (C) in the a) ^{1}Hα region and b) final amino protons region.



Supplementary 4. Infrared spectrum captured for zidovudine dissolved in HEPES.

5. 1. Influence of some biochemical variables on the insulin aggregation process and modelling of aggregation time

Insulin aggregation and association (see Scheme 5.1. a) and b)) are two natural processes that occur to insulin. They both lead to the formation of insoluble forms of insulin, but there are some differences between them. The secondary structure of the final forms is different: the native structure changes in aggregation, but remains unchanged in association [28, 41, 45, 46, 48, 68].

It is well-known that medium conditions can delay these processes or favour the occurrence of one over the other. Of all the biochemical variables, temperature (T) [13, 28, 69-71], ionic strength (I) [28, 72, 73] and acidity (pH) [46, 74-76] are the ones that have been subject to most study and which can be most readily found in the literature. The usual method for studying the influence of these variables on these processes is to vary the values of a single biochemical variable. This procedure ignores the synergic effect that the interaction of the variables can have on the processes.

This section discusses two aspects of the effect that some biochemical variables have on the processes. In the first part of this section, we assess the simultaneous influence of T, I and pH on the aggregation/association processes of insulin. The analytical method uses experimental design techniques and MCR-ALS analysis of the infrared spectra collected over time.

The second part of this section focuses on finding a response surface representative of the relationship between the aggregation time and the values of the biochemical variables. The same biochemical variables were taken into account and two different types of insulin were considered: one had zinc in the

solution while the other did not. Experimental design techniques were used to establish the response surface in both types of insulin. This issue may be of interest to the pharmaceutical industry, because insulin must be stored in conditions that do not allow insoluble components to form.

Again, a more comprehensive explanation of the experimental procedure of both studies is included as two research papers in the following section.

5.2.1. Results

MCR-ALS analysis of insulin aggregation / association processes. Influence of biochemical variables

Idoia Martí-Aluja and M. Soledad Larrechi

Chemometrics and Intelligent Laboratory Systems 127 (2013) 49 – 54

MCR-ALS analysis of insulin aggregation/association processes. Influence of biochemical variables

Abstract. The influence of medium conditions on insulin aggregation/association is of great interest from both a therapeutic and a pharmacological perspective. It is widely accepted that temperature, concentration of free protons (pH) and ionic strength play an important role in these processes, but studying these aspects in isolation ignores the interaction between variables. In this study, we used a chemometrics approach based on experimental design techniques and multivariate curve resolution-alternating least squares of the infrared spectra recorded in the association/aggregation process in order to simultaneously analyse the effect of acidity, temperature and ionic strength. The influence of acidity was studied at pH ranging between 3 and 7, temperature at between 35 ºC and 45 ºC and ionic strength at between 0.3 mM and 1.5 mM. The experiments were designed following a 2^3 full factorial design and the infrared spectra were collected between 1100 and 2000 cm^{-1}. An ANOVA analysis of the effect on the aggregation/association time revealed that the effect of temperature and pH cannot be quantified without also considering ionic strength. Concentration and spectral profiles recovered by means of multivariate curve resolution-alternating least squares allowed us to conclude that, regardless of temperature and ionic strength, the association process is promoted by neutral pH, while the aggregation process is promoted by acidic conditions. The uncertainty of the solutions was assessed by calculating the feasible band boundaries of the solutions. Optical images of each final product were captured to characterise the influence of the medium conditions on their morphology. The method we propose may be of great interest in many types of bioanalytical studies.

Keywords – Insulin; experimental design; insulin aggregation; insulin association; infrared spectroscopy (FTIR-ATR); MCR-ALS; Chemometrics

INTRODUCTION

Insulin is a protein that can adopt different conformations, including monomeric, dimeric and tetrameric [1-3]. These functional forms have a natural propensity to self-assemble. Several pathways have been described in the literature to explain this phenomenon [4], but they can be divided into two main processes: association and aggregation. Association is described as a reversible phenomenon that involves the interaction between two or more insulin monomers in their native secondary structure to form spherical arrangements with the same native structure. Insulin aggregation also involves the interaction of several previously unfolded insulin forms, resulting in aggregates that have lost the native structure of insulin [5-8]. Both processes, which might have important consequences for the bioactivity of insulin, are affected by the medium conditions [5,6]. From a pharmacological perspective [9,10], the study of the biochemical variables that affect insulin self-association and the evidence of the different pathways (aggregation or association) as a function of the medium conditions can help guide the process development and/or formulation effort [4,9]. From an analytical point of view, identifying whether the self-association may be related to an association or aggregation process may help design optimal chromatographic steps to remove aggregates disruption before their analysis [4].

Temperature (T), concentration of free protons (pH) and ionic strength (I) values play an important role in promoting the formation of aggregates and their morphology [11-19]. To our knowledge, studies conducted to date on these values consist of independent analyses of how each variable affects the process, without considering the possible interaction between them.

The aim of this study was to simultaneously assess the influence of temperature (T), pH and ionic strength (I) on insulin aggregation/association processes. The proposed method is based on experimental design techniques and multivariate curve resolution methods and allowed us to determine the effect of each individual variable as well as the effect of the interaction between them. It also provided useful information to evaluate the influence of these variables and their interactions on the pathway of these processes.

Eight experiments designed following a 2^3 full factorial design were performed. Infrared spectroscopy (IR) was the instrumental technique chosen to monitor the processes, as it has already proven useful in the study of aggregation [20,21]. The amide I, II and III regions are primarily used to assign secondary structures in protein processes. The amide I band (1600 - 1700 cm^{-1}) is mainly associated with the C=O stretching vibration and is directly related to the backbone conformation. Amide II (1500 - 1600 cm^{-1}) results from the N-H bending vibration and from the C-N stretching vibration. Amide III (1100 - 1330 cm^{-1}) analysis is used as a complementary method to amide I analysis in protein structural studies. The combination of the information contained in all of these regions provides insight into how a protein evolves [22-27].

We used the aggregation/association time provided by the scores plot for the first principal component obtained by means of a principal component analysis (PCA) to assess the effect of the biochemical variables [20]. The significance of the effects was evaluated using an ANOVA test. The number of significant insulin forms involved in each experiment was determined by means of singular value decomposition (SVD). The influence of medium conditions on the aggregation/association pathway was determined by means of multivariate curve resolution alternating least squares (MCR-ALS) [28,29]. The concentration and spectral profiles recovered by MCR-ALS enable us to detect which process

is most favoured by the medium conditions. The extent of uncertainty in the solutions calculated by means of MCR-ALS was evaluated using the MCR-BANDS method [30,31].The primary consequences of medium conditions on the morphology of the final forms were evaluated by means of optical imaging.

MATERIALS AND METHODS

Chemical reagents

Zn-free human insulin in a HEPES sodium salt buffer (pH=8.2) at a concentration of 10 mg/ml (1.72mM), sodium chloride (NaCl) and acetic acid (20%) were purchased from Sigma-Aldrich and used without further purification.

Experiments

Eight experiments were performed following a 2^3 full factorial experimental design; the experimental plan is depicted in Table 1a. The chosen biochemical variables and their high and low values were: temperature (35 ºC and 45 ºC), pH (3 and 7) and ionic strength (NaCl at 0.3 mM and 1.5 mM). In addition, a triplicate experiment was prepared under the experimental conditions corresponding to the central points of the experimental domain to evaluate the experimental error (Table 1b).

Experimental procedure and FTIR acquisition

The experimental procedure involved mixing the necessary amounts of insulin, acetic acid and NaCl for each experiment. Immediately after mixing, a drop of the mixture was placed on a small diamond crystal in the ATR cell of the spectrophotometer (FTIR 680 Plus JASCO and an RS232 Control) for measuring,

which was continuously purged with N_2. The FTIR spectra were recorded *in situ* every minute for 90 minutes in the spectral range of 1100 - 2000 cm^{-1}, every 0.964 cm^{-1}. The contribution of CO_2 was removed with the control software Spectra Manager before the spectra were exported into Matlab for further treatment [32]. The 90 infrared spectra from each experiment were arranged as rows in matrices $\mathbf{D_i}$, whose size were 90 x 934. Savitzky-Golay smoothing with a windows size of 7 and a third polynomial order was applied to all data sets in order to suppress the instrumental noise.

Table 1. (a) Experimental plan and aggregation time of each experiment; (b) experiments in the central points of the experimental domain.

a)

Experiment number	EXPERIMENTAL PLAN			Response
	Temperature (ºC)	*pH*	*Ionic strength I (mM)*	*Aggregation time (min)*
1	35	3	0.3	67
2	45	3	0.3	35
3	35	7	0.3	43
4	45	7	0.3	14
5	35	3	1.5	35
6	45	3	1.5	33
7	35	7	1.5	48
8	45	7	1.5	28

b)

Experiments in the central points of the experimental design				
Experiment number	*Temperature (ºC)*	*pH*	*Ionic strength I (mM)*	*Aggregation time (min)*
1	40	5	0.9	53
2	40	5	0.9	48
3	40	5	0.9	51

Chemometrics analysis and procedure

The influence of the biochemical variables on aggregation/association was analysed as follows:

- First, the transition time for each experiment was obtained from the scores plot for the first principal component [20]. These values were obtained when each individual data matrix ($\mathbf{D_i}$) was analysed by means of principal component analysis (PCA), according to the following equation: $\mathbf{D_i = TP^T}$, where **T** and $\mathbf{P^T}$ are the scores and loadings matrices.

- Second, the coefficient value representative of the effect of each biochemical variable (b_i) and their interactions ($b_{i,j,\ i\neq j}$) on the transition time were evaluated as the average of the transition times found in all of the experiments. The significance of these effects was evaluated by means of an ANOVA test using the variance of the responses on the central points as a measure of the experimental error.

- Third, the number of insulin forms involved in each experiment was assessed by means of the singular value decomposition (SVD) [33] of each individual data matrix, according to: $\mathbf{D_i = USV^T}$. The diagonal values of **S** are known as the *singular values* of $\mathbf{D_i}$, and the columns of **U** and the columns of $\mathbf{V^T}$ are the *left-singular vectors* and r*ight-singular vectors* of $\mathbf{D_i}$, respectively.

- Finally, the infrared spectral profiles of the insulin forms ($\mathbf{S^T}$) and their concentration profiles (**C**) in each experiment were estimated by multivariate curve resolution-alternating least squares (MCR-ALS). In this method, the data matrices are decomposed according to the following equation $\mathbf{D= CS^T+ E}$, where **E** contains the information not explained by the model. Once the number of significant factors needed to explain the chemical information of the data

matrices was chosen, initial estimates of $\mathbf{S}^T$ were built by means of evolving factor analysis (EFA) [28-31]. These initial estimates were iteratively optimised by a constrained alternating least squares (ALS) regression procedure. In the optimisation step of the ALS, the constraints applied to the process were non-negativity and closure for the concentration profiles and non-negativity and normalisation for the spectral profiles [28,29]. The quality of the MCR-ALS was evaluated using the lack of fit (LOF) and the percentage of explained variance (R^2) parameters, which identify the dissimilarity between the experimental data matrix (**D**) and the reproduced by MCR-ALS. The extent of uncertainty in the estimated solutions was evaluated by calculating the feasible bands around each individual solution by means of MCR-BANDS algorithm [29-31].

PCA, MCR-ALS and MCR-BANDS analysis were made under Matlab environment [32]. Calculations for the experimental design were made using STATGRAPHICS Plus 5.0 [34].

RESULTS AND DISCUSSION

The spectra recorded in all experiments follow two different patterns; experiments 3 and 5 are representative of each situation. To facilitate the discussion, only the initial (solid line) and final (dotted line) spectra from these experiments are shown in Fig. 1a. The spectra have been shifted up the y-axis to clarify the discussion. The initial spectrum of both experiments present the marker for the α-helix structure of native insulin in the amide I region, located at around 1636 cm^{-1} [27, 35-38]. In the spectrum of experiments at pH 3 (experiment 5) a less intense absorption band at around 1280 cm^{-1} can also be seen in the amide III region, representative of random structures, in which there is a random distribution of the backbone coils of the insulin. The spectra

exhibit a shift over time of the maxima in the amide I region from 1636 cm^{-1} to 1647 cm^{-1}, which is indicative of aggregates formation. Moreover, new bands in the amide II region emerged around 1530 cm^{-1}, which are also representative of the β-sheet structure of aggregates [22-24].

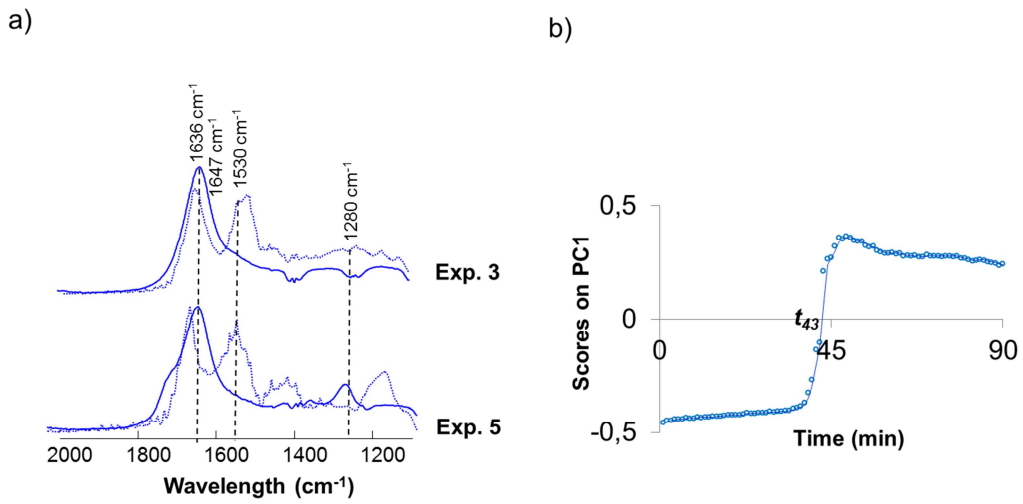


Figure 1. (a) The initial (solid line) and final (dotted line) spectra of experiment 3 and 5. (b) Scores plot for the first principal component of experiment 3.

Aggregation and association processes imply a change in state in which insulin evolves from a solution to form an insoluble solid, which represents the main variability that takes place during the process. A measure to detect the time at which the state change occurs can be obtained by inspecting the scores plot for the first principal component, calculated by means of PCA [20]. To facilitate the discussion, Figure 1b shows the scores on the first principal component (PC) for experiment 3, which contained 99% of the variance. The transition time of each experiment provided by the PCA of the individual matrices Di are indicated in the last column of Table 1a. These values were used as responses to quantify

the effect of the biochemical variables on the aggregation/association processes.

Table 2. Results of ANOVA test for the relevant biochemical variables.

Estimated effects		Results of the ANOVA test			
		Sum squares	*d.f.*	*Mean square*	*F ratio*
b_0: **average**	37.8				
b_1: **T**	-20.7	861.1	1	861.1	**272.5**
b_2: **pH**	-9.3	171.1	1	171.1	**54.2**
b_3: **I**	-3.8	28.1	1	28.1	8.9
b_{12}: **T x pH**	-3.8	28.1	1	28.1	8.9
b_{13}: **T x I**	9.8	190.1	1	190.1	**60.2**
b_{23}: **pH x I**	13.3	351.1	1	351.1	**111.1**
Mean square of experimental error = 3.16					
$F_{critic\ (1,2,\ \alpha=\ 0.05\ )}$ **= 18.5**					

Table 2 shows the coefficients of the effect on the transition time of the individual biochemical variables (T, pH and ionic strength), the interaction between them and the results of the ANOVA test. According to the F-ratios calculated at a significance level of α= 0.05, the effect of T, pH and the interaction between them and ionic strength significantly differed from the experimental error ($F_{effect} > F_{critic}$). Therefore, the transition time varied considerably when T or pH changed. Although the effect of ionic strength on the studied range might not be considered significant, its interaction with T and pH is noteworthy as it is reflected in their coefficient values. Because the effect of the interaction between variables is quite pronounced, a deeper analysis of the variables is called for. An easy way to achieve this is by analysing the effect plots. By way of example, Figure 2 shows the effect plot of the interaction

between T and ionic strength at both pH values. In this figure, the transition time of every experiment for each pH is included in the middle of each square. The interaction between T and I can be evaluated by the difference (Δ) between the transition time of the experiments at low and high T when the pH is fixed (Figure 2a and 2b). In both pH values, the transition time decreased when the experiments were carried out at 45 ºC. The difference value (Δ) was higher in the experiments at the lowest value of ionic strength (Δ= - 32 at pH 3 and Δ= -29 at pH 7), compared to those at high values (Δ= - 2 at pH 3 and Δ= -20 at pH 7). At this point, we can state with some confidence that individual analyses can hide the real effect of the variables in the process.

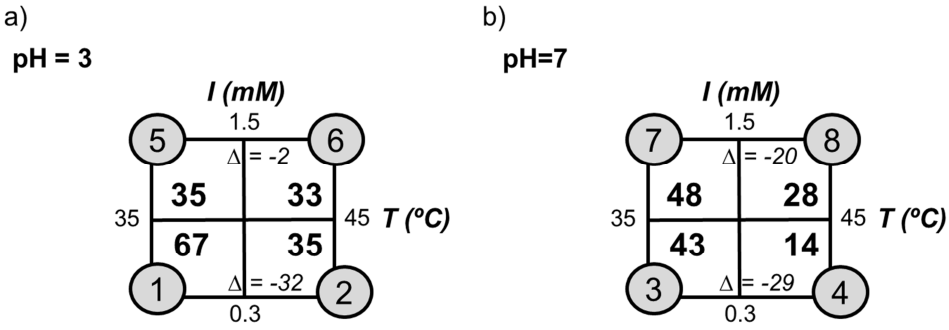


Figure 2. Effect plot of the interaction between temperature and ionic strength at a) pH 3 and b) pH 7.

To obtain information about the influence of medium conditions in the pathway of the process, the number of insulin forms present in each experiment was obtained by the SVD analysis of each individual data matrix ($\mathbf{D_i}$) [33]. The first six *singular values* obtained from the individual SVD analyses of each experimental matrix ($\mathbf{D_1}$-$\mathbf{D_8}$) and their explained variance (%) (in brackets) are depicted in Table S-1 (supplementary material). The *right-singular vectors* (**V**) associated with each *singular value* were checked to rule out the possibility

that the excluded *singular values* contain chemical information [33]. In all cases, it was observed that factors with low variance retain chemical information. Three significant values were retained in the experiments performed at pH 7 (experiments 3, 4, 7 and 8) and four were found in the experiments at pH 3 (experiments 1, 2, 5 and 6). 99.9% of the explained variance was retained in all cases.

MCR-ALS is a useful tool for revealing the chemical information represented by factors with low variance because it solves the data matrix by incorporating additional information through restrictions with chemical meaning. MCR-ALS was calculated for all individual matrices using a convergence criterion of 0.1%. Lack of fit and explained variance were also calculated and in all cases the values below 1.8% and equal to or above 99%, respectively.

Again, we will only discuss the results of one experiment for each situation (experiments 3 and 5). Fig. 3a shows the spectra recovered for the three insulin forms in experiment 3 at pH 7. The spectral features of the α-helix structure in the amide I region (1636 cm^{-1}) are present in the recovered spectra of all three forms. The third and fourth spectra show absorption bands at 1515 cm^{-1} and 1545 cm^{-1}, representative of β-sheet and β-turn structures.

The concentration profiles obtained for experiment 3 (pH 7) are represented in arbitrary units in Fig. 3b. These profiles contain information about the evolution of each insulin form involved in the process. Although two forms appear before the transition time (43 min), the concentration of the native form (first form) remains virtually the same. The second form appears relatively close to the transition time. After this point, a change in physical state takes place and only one type of form is present. This behaviour is consistent with the association process described in the literature. In this process, two or more monomers

interact electrostatically without losing the α-helix structure of native insulin [5,6]. At a pH higher than the isoelectric point (pI 5.5), insulin is negatively charged and these electrostatic interactions stabilise the solution and make the assembly process energetically favourable. The associated forms can evolve into spherical precipitates, which can be characterised by optical microscopy. The optical image captured at the end of the process (Fig. 4a) indicates that the insulin molecules are bound together, forming a grouped and globular dense material. The results obtained are in agreement with those reported in the literature [11,16,44]; therefore, these medium conditions favour the association of insulin [15, 40-43]. This association may take place faster depending on the other biochemical variables assessed in this study.

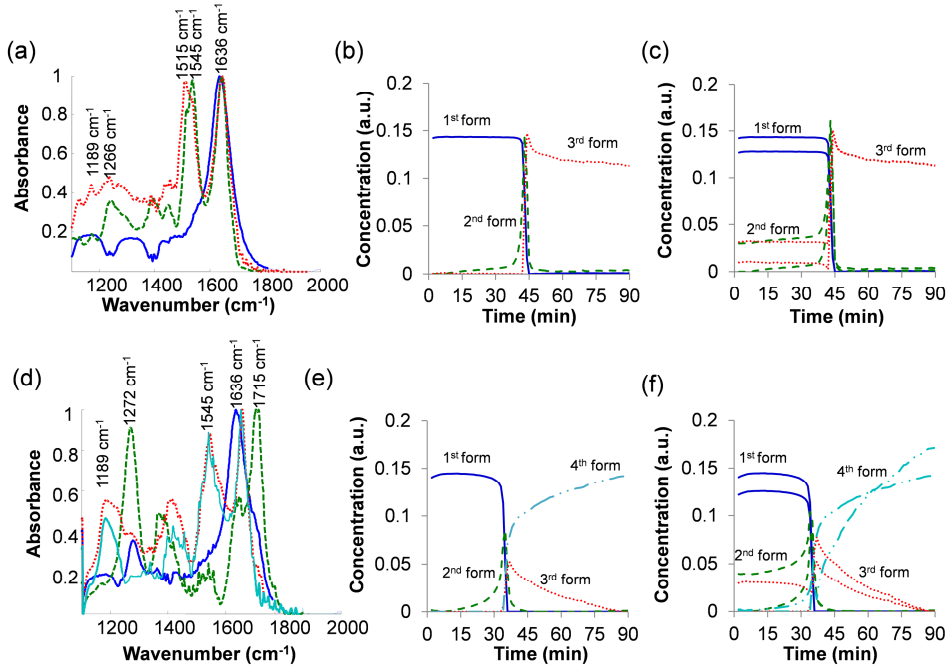


Figure 3. Experiment 3: (a) spectral profiles, (b) concentration profiles and (c) feasible bands for the concentration profiles. Experiment 5: (d) spectral profiles, (e) concentration profiles and (f) feasible bands for the concentration profiles (solid line: first forms, dashed line: second forms, dotted line: third forms and dashed/dotted line: fourth forms).

The recovered spectra obtained for experiment 5 are shown in Figure 3d. The first spectrum presents the same bands as in experiment 3 (Fig. 3c, α-helix structures). The second spectrum is mainly composed of the spectral bands representative of random-coils (1272 cm^{-1}) and β-turn structures (1715 cm^{-1}). These considerations suggest that the second forms differ in some way from those observed in experiment 3 (Fig. 3a). The spectra of the third and fourth forms have the spectral features of β-sheet structures (1189 cm^{-1}, 1545 cm^{-1} and 1660 cm^{-1}), although significant differences were present in the amide III region (1100 - 1330 cm^{-1}), which may indicate that the two forms present different secondary structure compositions.

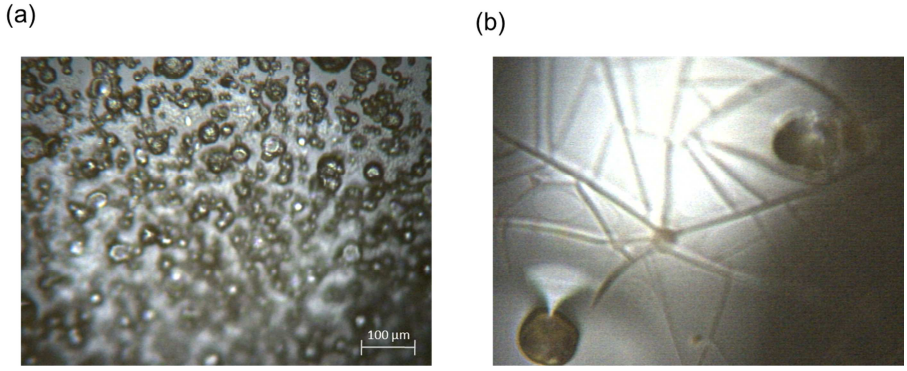


Figure 4. Optical images at the end of the process of (a) experiment 3 and (b) experiment 5.

The evolution of the concentration profiles for the four forms at pH 3 is shown in Figure 3e. The concentration of insulin in its native form decreases before the transition time (35 minutes), while the concentration of the second forms enriched with β-structures increases. After this time, these second forms are totally consumed. The evolution of the concentrations of the other two forms that appear after the transition point may be representative of a process in which a transition takes place between two forms, more or less enriched with β-structures. Globally speaking, these results may be representative of the

insulin aggregation process. In that process, the initial forms of insulin slowly lose their native structure over a period of 35 minutes during the unfolding process. After a significant level of unfolding forms has been reached, the forms in which random-coil and β-turn structures dominate (second forms at 35 minutes), they bind together forming nucleus forms (third forms), which grow into ordered aggregates (fourth forms) [5-8,10]. A complementary understanding of this process was obtained by optical imaging. The optical image captured at the end of the process (Fig. 4b) shows thin and ordered aggregates with smaller diameters. Aggregation under these conditions seems to be sequential and ordered.

As there is rotational ambiguity in MCR-ALS solutions, the feasible band boundaries related to each of the solutions were calculated using the MCR-BANDS method. The values of f_n^{min} and f_n^{max} and the difference between them are given for every component of each experiment in Table S-2 (supplementary material). The extent of rotational ambiguity was assessed using the difference value: the closer to zero the value is, the less rotational ambiguity the solution will present. Overall, we can see that most of the difference values are in the range of 0 to 0.20, which means that the rotational ambiguity cannot be considered significant. In experiment 8 (pH 7, T= 45 ºC and I = 1.5 mM), this value is higher (0.67 and 0.74, respectively, for the second and third forms), which means that the extent of rotation ambiguities was also high. A possible explanation might be that these are the most favourable conditions for electrostatic interactions.

As the spectral profiles of the different compounds do not differ significantly, only the f_n^{min} and f_n^{max} values of the concentration profiles are shown in Fig. 3c (experiment 3, pH 7) and Fig. 3f (experiment 5). The evolution of the concentration profiles at pH 7(Fig. 3c) indicates that three forms are present

from the beginning. Our results are unclear basically in the first stage of the process (before the transition point) and therefore this ambiguity does not affect the conclusions drawn above. Although at pH 3 (Fig. 3f) nucleus forms (third forms) are present before the transition time, they emerge near the transition time, so they do not significantly affect our conclusions. Overall, the uncertainty related to the results does not affect our final conclusions for either process.

CONCLUSIONS

The conjunction of experimental design techniques and MCR-ALS has proved a useful tool for the analysis of processes involving insulin. The influences of temperature (T), pH and ionic strength (I) in insulin aggregation/association cannot be analysed individually. T and pH were the most influential variables, but they both interact with ionic strength in a significant way.

Analysis of infrared data using the MCR-ALS method has proved a powerful technique to demonstrate the influence of medium conditions on these processes. The concentration and spectral profiles recovered for experiments at pH 7 allow us to postulate the existence of association process, independently of T and I. Moreover, these profiles allow us to suggest that the aggregation process took place in experiments at pH 3. All final forms were characterised by optical imaging.

The extent of uncertainty in the solutions recovered by MCR-ALS was assessed by means of the MCR-BANDS method to provide a more accurate interpretation of our results.

ACKNOWLEDGMENTS. We would like to thank Universitat Rovira i Virgili for providing Idoia Martí-Aluja with a doctoral fellowship.

REFERENCES

[1] D-P. Hong, A. Ahmad, A. L. Fink, *Biochemistry* **45** (2006) 9342 – 9353.

[2] Y. Hong, L. Meng, S. Chen, C. W. T. Leung, L-T. Da, M. Faisal, D-A. Silvia, J. Liu, J. W. Y. Lam, X. Huang, B. Z. Tang, *J. Am. Chem. Soc*. **134** (2012) 1680 – 1689.

[3] E. J. Nettleton, P. Tito, M. Sunde, M. Bouchard, C. M. Dobson, C.V. Robinson, *Biophys. J.* **79** (2000) 1053 – 1065.

[4] J. S. Philo, T. Arakawa, *Curr. Pharm. Biotech*. **10** (2009) 348 – 351.

[5] J. Brange, L. Andersen, E. D. Laursen, G. Meyn, E. Rasmussen, *J. Pharm. Sci.* **86** (1997) 517 – 525.

[6] Q. Hua, M. A. Weiss, *J. Biol. Chem*. **279** (2004) 21449 – 21460.

[7] A. Ahmad, V. N. Uversky, D. Hong, A. L. Fink, *J. Biol. Chem.* **280** (2005) 42669 – 42675.

[8] K. Giger, R. P. Vanam, E. Seyrek, P. L. Dublin, *Biomacromolecules* **9** (2008) 2338 – 2344.

[9] D. N. Brems, L. A. Alter, M.J. Beckage, R. E. Chance, R. D. DiMarchi, L. K. Green, H. B. Long, A. H. Pekar, J. E. Shields, B. H. Frank, *Protein Eng.* **5** (1992) 527 – 533.

[10] J. Brange, L. Langkjoer, *Pharm. Biotechnol*. **5** (1993) 315 – 350.

[11] V. Smirnovas, R. Winter, *Biophys. J.* **94** (2008) 3241 – 3246.

[12] T. J. Gibson, R. M. Murphy, *Protein Sci.* **15** (2006) 1133 – 1141.

[13] M. I. Dyukov, M. P. Grudinin, A. K. Sirotkin, O. I. Kiselev, *Biochemistry Biophys*. **419** (2008) 79 – 81.

[14] A. Gerhardt, K. Bonam, J. S. Bee, J. F. Carpenter, T. W. Randolph, *J. Pharm. Sci.* **102** (2013) 429 – 440.

[15] A. K. Attri, C. Fernádez, A. P. Minton, *Biophys. Chem.* **148** (2010) 23 – 27.

[16] S. L. Shammas, T. P. J. Knowles, A. J. Baldwin, C. E. MacPhee, M. E. Welland, C. M. Dobson, G. L. Devlin, *Biophys. J.* **100** (2011) 2783 – 2791.

[17] A. Ahmad, I. S. Millett, S. Doniach, V. N. Uversky, A. L. Fink, *Biochemistry* **39** (2003) 11404 – 11416.

[18] K. Huus, S. Havelund, H. B. Olsen, M. van de Weert, S. Frokjaer, *Biochemistry* **44** (2005) 11171 – 11177.

[19] C. Chandler, C. M. Gryniewicz, T. Pringle, F. Cunningham, *Am. J. Health. Syst. Pharm.* **10** (2008) 953 – 956.

[20] I. Martí-Aluja, J. Ferré, I. Ruisánchez, M. S. Larrechi, *Chemom. Intell. Lab. Syst.* **118** (2012) 180-186.

[21] I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi, *Anal. Chim. Acta* **760** (2013) 16 –24.

[22] G. Vedantham, H. G. Sparks, S. U. Sane, S. Tzannis, T, M. Przybycien, *Anal. Biochemistry* **1** (2000) 33 –49.

[23] P. Kupser, K. Pagel, J. Oomens, N. Polfer, B. Koksch, G. Meijer, G. von Helden, *J. Am. Chem. Soc*. **132** (2010) 2085 – 2093.

[24] L. P. DeFlores, Z. Ganim, R. A. Nicodemus, A. Tokmakoff, *J. Am. Chem. Soc*. **131** (2009) 3385 –3391.

[25] S. Cai, B. R. Singh, *Biophys. Chem.* **80** (1999) 7 – 20.

[26] M. Carbonaro, A. Nucara, *Amino Acids* **3** (2010) 679 – 690.

[27] F. H. Carpenter, *Am. J. Med*. **40** (1966) 750 – 758.

[28] C. Ruckebusch, L. Blanchet, *Anal. Chim. Acta* **765** (2013) 28 – 36.

[29] R. Tauler, *Chemom. Intell. Lab. Syst*. **30** (1995) 133 – 146.

[30] J. Jaumot, R. Tauler, *Chemom. Intell. Lab. Syst.* **103** (2010) 96 – 107.

[31] X. Zhang, R. Tauler, *Anal. Chim. Acta* **762** (2013) 25 – 38.

[32] The Mathworks, MATLAB Version 7.0, Natick, MA, 2004.

[33] D. L. Massart, B. Vandeginste, L. Buydens, S. de Jong, P. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A,* Elsevier, Amsterdam, 1997.

[34] Statgraphics <http://www.statgraphics.com/>

[35] A. Elliot, E. J. Ambrose, *Nature* **165** (1950) 921 – 922.

[36] S. Krimm, J. Bandekar, *Adv. Prot. Chem.* **38** (1986) 181 – 364.

[37] M. Bouchard, J. Zurdo, E. J. Nettleton, C. M. Dobson, C. V. Robinson, *Protein Sci.* **9** (2000) 1960 – 1967.

[38] K. Murayama, M. Tomida, *Biochemistry* **43** (2004) 11526 – 11532.

[39] N. M. Faber, *Anal. Chim. Acta* **439** (2001) 193 – 201.

[40] P. D. Jeffrey, B. K. Milthorpe, L. W. Nichol, *Biochemistry* **15** (1976) 4660 – 4665.

[41] A. K. Attri, C. Fernández, A. P. Minton, *Biophys. Chem.* **148** (2010) 28 – 33.

[42] M. E. M. Cromwell, E. Hilario, F. Jacobson, *AAPS J*. **8** (2006) E572 – E579.

[43] E. Y. Chi, S. Krishnan, T. W. Randolph, J. F. Carpenter, *Pharm. Res.* **20** (2003) 1325 – 1336.

[44] E. van der Linden, P. Venema, *Curr. Opin. Colloid In*. **12** (2007) 158 –165.

Supplementary material

Table S-1. First six *eigenvalues* obtained on the individual SVD analysis of each experimental matrix ($\mathbf{D_1}$-$\mathbf{D_8}$)

Factor	Experiment number							
	1	***2***	***3***	***4***	***5***	***6***	***7***	***8***
1	**21,54 (96,5%)**	**10,58 (96,4%)**	**14,40 (82,7%)**	**36,18 (99,1%)**	**14,08 (90,5%)**	**18,01 (85,4%)**	**19,82 (97,7%)**	**28,70 (99,4%)**
2	**5,88 (99,6%)**	**2,57 (99,5%)**	**3,45 (99,8%)**	**2,92 (99,8%)**	**3,54 (99,4%)**	**4,43 (99,3%)**	**3,42 (99,6%)**	**3,14 (99,9%)**
3	**2,23 (99,8%)**	**0,82 (99,8%)**	**0,51 (99,9%)**	**0,78 (99,9%)**	**0,92 (99,8%)**	**1,51 (99,7%)**	**0,32 (99,9%)**	**0,46 (99,9%)**
4	**0,58 (99,9%)**	**0,35 (99,9%)**	0,13 (99,9%)	0,24 (99,9%)	**0,32 (99,9%)**	**0,33 (99,8%)**	0,16 (99,9%)	0,23 (100%)
5	0,23 (100%)	0,12 (99,9%)	0,08 (99,9%)	0,14 (99,9%)	0,11 (99,9%)	0,24 (99,9%)	0,07 (99,9%)	0,12 (100%)
6	0,16 (100%)	0,08 (100%)	0,02 (100%)	0,12 (100%)	0,06 (99,9%)	0,18 (99,9%)	0,03 (99,9%)	0,04 (100%)

Table S-2. Values of f_n^{min} and f_n^{max} and the difference between for every component of each experiment.

	f	Form 1	Form 2	Form 3	Form 4
Exp. 1	fmax	0,44	0,20	0,34	0,66
	fmin	0,39	0,15	0,25	0,63
	fmax-fmin	**0,06**	**0,04**	**0,08**	**0,04**
Exp. 2	fmax	0,77	0,50	0,08	0,50
	fmin	0,39	0,12	0,04	0,44
	fmax-fmin	**0,38**	**0,38**	**0,03**	**0,06**
Exp. 3	fmax	0,58	0,24	0,76	-
	fmin	0,44	0,16	0,72	-
	fmax-fmin	**0,14**	**0,08**	**0,04**	-
Exp. 4	fmax	0,13	0,31	0,87	-
	fmin	0,11	0,28	0,83	-
	fmax-fmin	**0,02**	**0,03**	**0,04**	-
Exp. 5	fmax	0,56	0,24	0,37	0,69
	fmin	0,36	0,11	0,15	0,58
	fmax-fmin	**0,20**	**0,14**	**0,22**	**0,10**
Exp. 6	fmax	0,40	0,27	0,41	0,61
	fmin	0,31	0,19	0,38	0,59
	fmax-fmin	**0,08**	**0,08**	**0,03**	**0,02**
Exp. 7	fmax	0,87	0,32	0,42	-
	fmin	0,72	0,08	0,32	-
	fmax-fmin	**0,16**	**0,25**	**0,10**	-
Exp. 8	fmax	0,25	0,81	0,92	-
	fmin	0,18	0,14	0,18	-
	fmax-fmin	**0,08**	**0,67**	**0,74**	-

5.2.2. Results

Modelling insulin aggregation time using response surface methodology

Idoia Martí-Aluja and M. Soledad Larrechi

Submitted to *Chemometrics and Intelligent Laboratory Systems*

Modelling insulin aggregation time using response surface methodology

Abstract. Aggregation time of insulin was modelled in function of temperature, pH and ionic strength of the solution. Monomeric and hexameric insulin solutions were considered. The experiments were performed using a face centred central composite design. The experimental domain was defined by the following individual ranges: 35ºC and 45ºC for temperature, 3 and 7 for pH and a concentration of 0.3 and 0.6 mM sodium chloride, for ionic strength. The experiments were monitored by infrared spectroscopy and the aggregation time used as a response was obtained from the scores plot of the first principal component. The absence of bias was evaluated by a joint confidence interval test from the regression line obtained from the measured and predicted response (aggregation time). The standard deviation of the differences between the measured and the predicted results were comparable to the standard deviation of the experimental error. At human body temperature, aggregates take longer to form in the monomeric insulin solution at low values of ionic strength (I=0.20-0.33 mM) and acid pH (from 3-3.8). In the hexameric insulin solution they form more slowly at high pH (6.5-8) and at any ionic strength. The morphology of the aggregates formed was shown by optical microscopy.

Keywords - Insulin; experimental design; insulin aggregation; infrared spectroscopy (FTIR-ATR); response surface methodology; Chemometrics

INTRODUCTION

Insulin has been studied ever since it was first isolated in 1922 [1,2]. Insulin is a small helical protein hormone consisting of two polypeptide residues linked by two interchain disulphide bridges [3,4,5]. It can exist in solution as monomers, dimers, tetramers, hexamers and higher associated states, depending on the concentration, pH, metal ions, ionic strength, solvent composition, etc. [6,7,8].

It is well known for its propensity to self-associate to form fibrils or aggregates [2,9]. Several authors have proposed that insulin aggregation takes place because oligomers (dimers, tetramers, hexamers, etc.) dissociate into monomers, which then undergo a structural change [4,10,11,12,13,14]. Aggregated forms of insulin lose their activity and are not therapeutically effective, so several studies have been made to analyse the effect of the factors that promote or inhibit insulin aggregation. Studies on this process have shown that acidic pH, high temperature, ionic strength or agitation promote the formation of aggregates, and that the rate of aggregation depends on the type of acid used [15,16,17].

Many spectroscopic and microscopic techniques have been used to study aggregation processes, including fluorescence spectroscopy [18,19,20], circular dichroism (CD) [21,22,23,24], Raman spectroscopy [25,26,27] and scanning electron microscopy (SEM) [28,29]. In general, these studies focus on the analysis of the structural changes involved in this process and in the morphology of the aggregates. Information about the optimal conditions that delay or inhibit the formation of aggregates is taken from the comparisons made of the aggregation process carried out at different values of each individual biochemical variable [7,30, 31,32].

The literature review indicates that of all the biochemical variables, temperature is probably the most critical in the aggregation of insulin. A rise in temperature increases the rate of aggregation because hydrophobic interactions increase [30, 31,33, 34,35], which changes the relative composition of secondary structures and alters the aggregation behaviour. The second critical variable is the solution pH. The solution pH can strongly affect the rate of insulin aggregation by changing the type and distribution of charges in insulin [7, 11,15,33,35]. The ionic strength of the solution is another variable that has

been widely studied [7, 11,24,36]. All these conclusions are drawn from an individual analysis of the effect of the biochemical variables on the formation of aggregates. As far as we know, only the work recently published by us simultaneously analyses the effect of temperature, pH and ionic strength on the kinetics of the formation of insulin aggregates [37].

From a pharmacological and clinical point of view, it is important to know the relationship between temperature, pH and ionic strength with insulin aggregation time. The objective of the present study is to establish a predictive model to simultaneously determine the optimal conditions of these variables in which the formation of aggregates in an insulin solution is less favoured. The used strategy is based on experimental design techniques which enable us to find this relationship using a small amount of experiments.

The formation of the aggregates was optimised for two different types of insulin: hexameric (which contains Zn in the solution) and monomeric. The interest in comparing two different types of insulin lies in the fact that the hexameric form is the most stable whereas the active and reactive is the monomeric. Each predictive model was found using response surface methodology (RSM) [38]. Eighteen experiments were performed in the experimental conditions defined by two face centred central composite design (CCD), which were used to establish the model of aggregation time of insulin. The experimental domain was defined by the following individual ranges: 35ºC and 45ºC for temperature, 3 and 7 for pH and a concentration of 0.3 and 0.6 mM sodium chloride, for ionic strength.

All experiments were monitored by infrared spectroscopy (IR) and the aggregation time provided by the scores plot for the first principal component

obtained by means of principal components analysis (PCA) was used as a response [37,39,40].

The models were validated by testing the aggregation time predicted in eight new experiments with each one of the insulin samples. The joint confidence interval test for the individual regression line obtained from the experimental and predicted time aggregation was used to assess their trueness. The standard deviation of the differences between the measured and the predicted results (RMSEP) was 1.96 and 2.66 for the monomeric and the hexameric insulin, respectively. Both values are comparable to the standard deviation representative of the experimental error.

In addition, optical microscopy was used to study the morphology of the aggregates formed and establish how these are affected by the biochemical variables.

MATERIALS AND METHODS

Insulin samples and chemical reagents

Two different types of insulin were used: monomeric and hexameric. In the first case, it was a Zn-free human insulin in a HEPES sodium salt buffer (pH=8.2) at a concentration of 10 mg/ml (1.72mM) (Sigma-Aldrich) and, in the second, a Zn-insulin (purchased in a commercial chemist).

Sodium chloride (NaCl) and acetic acid were purchased from Sigma-Aldrich and used without further purification.

Experimental procedure and FTIR acquisition

The experimental procedure involved mixing the necessary amounts of insulin, acetic acid and NaCl for each experiment. Immediately after mixing, a drop of the mixture was placed for measurement on a small diamond crystal in the ATR cell of the spectrophotometer (FTIR 680 Plus JASCO and an RS232 Control), which was continuously purged with N_2. The FTIR spectra were recorded *in situ* every minute for 90 minutes in the spectral range 1100 - 2000 cm^{-1}, every 0.964 cm^{-1}. The contribution of CO_2 was removed with the control software Spectra Manager before the spectra were exported into Matlab for further treatment [41]. The 90 infrared spectra from each experiment were arranged as rows in matrices $\mathbf{D}_i$ (90 x 934).

The data were processed under Matlab environment [41]. Calculations for the experimental design were made using STATGRAPHICS Plus 5.0 [42].

Establishment of the predictive model

The dependence between the time that insulin takes to aggregate and the values of temperature, pH and ionic strength is studied using experimental design techniques. Insulin aggregation was studied in the experimental domain defined by the following individual ranges: 35ºC and 45ºC for temperature, 3 and 7 for pH and a concentration of 0.3 and 0.6 mM sodium chloride, for ionic strength. In terms of experimental design, the higher and lower values of each biochemical variable are respectively codified as (+1) and (-1). The response variable considered as representative of insulin aggregation was the time at which the sample changes from a liquid to a solid state (aggregation time, t_m).

Response surface methodology (RSM) was used to establish the model that predicts the aggregation time of insulin from the values of the biochemical variables.

First, a first-order model was established, according to:

$$t_m = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \qquad (1)$$

where t_m is the response value, b_i are the linear coefficients which describe the sensitivity of the response to the variations in the corresponding variables (x_1, x_2 and x_3 refer to codified temperature, pH and ionic strength, respectively), and b_{ij} are the crossproduct coefficients used to establish the first-order response surface [38].

The first eight experiments in Table 1 were used to calculate the coefficients (b_i) which correspond to a 2^3 full factorial design.

To validate this model, we evaluated the sum of squares for pure quadratic curvature ($SS_{pure\ quadratic}$):

$$SS_{pure\ quadratic} = \frac{n_F n_C \left(\overline{t_{m,F}} - \overline{t_{m,C}}\right)^2}{n_F + n_C} \qquad (2)$$

where $\overline{t_{m,F}}$ is the average of the eight experiments in the experimental plan, $\overline{t_{m,C}}$ the average of the experimental responses at the central point, n_F the number of points in the factorial design and n_C the number of replicates at the central point.

The sum of squares for pure quadratic curvature ($SS_{pure\ quadratic}$) was compared with the $SS_{residual}$ obtained with the replicates of the central point by means of an F-test. There is no quadratic curvature when both values are similar. In this study, this test did not pass, so that a second-order model was postulated:

$$t_m = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \quad (3)$$

where t_m, b_0, b_i, b_{ij}, x_i and x_j have the same meaning as in the previous case and b_{ii} is a new coefficient that corresponds to the quadratic terms that describe the curvature [38].

A common strategy for obtaining second-order response surfaces is to use face centred central composite design (CCD). This design locates the axial points on the centres of the faces of the cube from the full factorial design 2^3. In the present work, two CCD (18 runs: 2^3 + 4 centre points + 6 star points) were performed to calculate the relation between the insulin aggregation time and the values of the biochemical variables considered. The experimental conditions that correspond to the star points of the CCD have been included at the end of Table 1. The suitability of the experimental data to the second-order model was checked by the analysis of the residual plot of the experiments.

Table 1. Experimental plan and aggregation time for each experiment for monomeric and hexameric insulin.

Experiment number	Temperature (ºC)	pH	Ionic strength (mM)	*Aggregation time (min)*	
				Monomeric insulin	Hexameric insulin
1	35	3	0.30	67	30
2	45	3	0.30	35	13
3	35	7	0.30	43	46
4	45	7	0.30	14	41
5	35	3	0.60	38	35
6	45	3	0.60	12	38
7	35	7	0.60	29	74
8	45	7	0.60	23	42
PC1	40	5	0.45	23	29
PC2	40	5	0.45	24	25
PC3	40	5	0.45	25	24
PC4	40	5	0.45	25	26
CC1	32	5	0.45	53	57
CC2	48	5	0.45	14	35
CC3	40	2	0.45	21	18
CC4	40	8	0.45	30	25
CC5	40	5	0.20	32	24
CC6	40	5	0.70	28	24

The absence of bias in the models was assessed by performing a join statistical test of the slope and intercept. Other eight experiments within the experimental domain were performed for each type of insulin. The results predicted for each model ($t_{m,i,predicted,}$) were regressed against its measured ($t_{m,i,measured}$) values. If there are no significance differences between the two values (measured and predicted), the theoretical point (0,1) falls within the joint confidence interval of the ellipse defined around the slope (b_0) and intercept (b_1) of the straight regression line [43].

The accuracy in the aggregation time was estimated in terms of the value of the root-mean-square error of prediction (RMSEP) of the new experiments.

RESULTS AND DISCUSSION

To facilitate the discussion, Figure 1 shows only the spectra recorded during aggregation in the experimental conditions of experiment 8 (see Table 1) for monomeric insulin (Fig. 1a) and hexameric insulin (Fig. 1b).

The particular ability of infrared spectroscopy to detect the structural changes involved in the insulin aggregation process has been extensively discussed in numerous studies [37, 39,40,44]. For this reason, the spectral discussion in this study focusses on the kinetic evolution of the spectral bands in the amide I region (from 1636 cm^{-1} to 1653 cm^{-1}) that reflect the structural change from an α-helix structure to a β-sheet indicative of the aggregate formation. The evolution of this spectral band is similar in both figures, although the kinetic variations are more pronounced for monomeric insulin (Fig. 1a). While the monomeric insulin shows an abrupt change over time, a more progressive evolution is detected in hexameric insulin. According to literature there is a lag time, which marks the beginning of the growth of the aggregates. One possible explanation for the behaviour observed in monomeric insulin may be that the species formed before this time (Fig. 1a) evolve immediately to final aggregates, which may explain the abrupt evolution. On the other hand, their evolution in hexameric insulin is more progressive and the conversion is slower (Fig. 1b) [7, 11,45]. Additionally, various mechanisms have been reported in the literature to explain the formation of the aggregates, but they all describe an alteration that occurs in the monomeric form [46,47,48,49]. Therefore, one explanation of the kinetic behaviour shown in Figure 1 is that the hexamers of hexameric insulin are first converted into monomers, which makes the process slower (Fig. 1b).

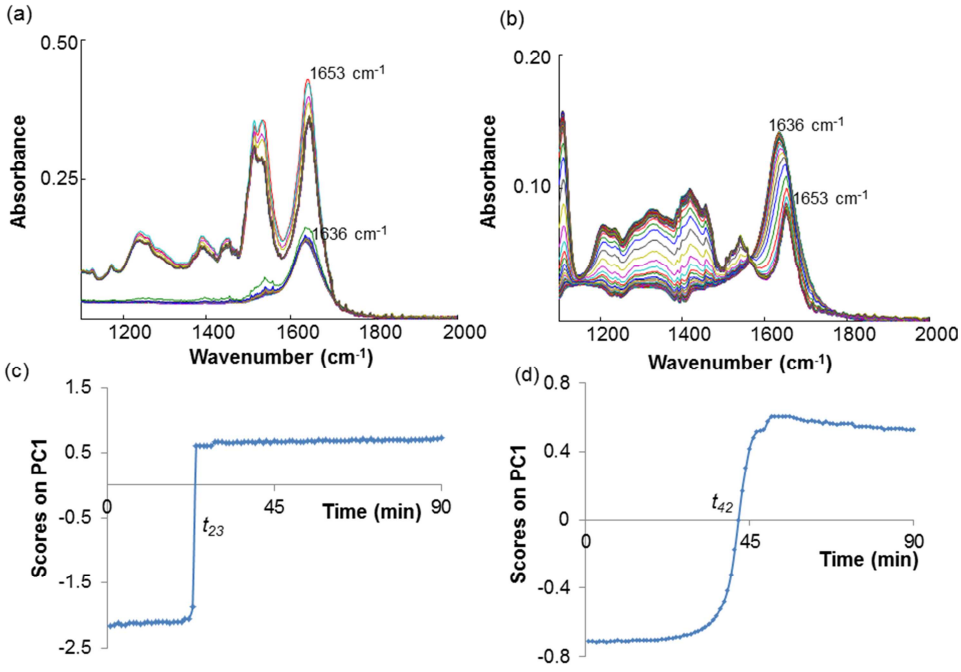


Figure 1. Infrared spectra acquired every minute of experiment 8 for (a) monomeric and (b) hexameric insulin. Scores plot on the first principal component for (c) monomeric and (d) hexameric insulin.

Although the evolution of the intensities of this spectral band provides information about the kinetics of the process, it is difficult to find a systematic pattern over time. The variations are complex because they are dependent on the variations of other bands that undergo minor changes. In a previous study, we demonstrated the advantages of subjecting the spectra recorded over time to principal components analysis (PCA) [37,39,40]. PCA analysis is calculated after the data have been mean-centred and the common part of the time-series spectra removed, thus allowing us to focus on the spectral changes around the average spectrum. Figure 1c and 1d shows the scores plot for the first principal component calculated with experiment 8 for monomeric and hexameric insulin, respectively. The profile of this plot is representative of the

physical and chemical changes that occur during insulin aggregation. The inflexion point marks the transition time between the liquid state of native insulin and the solid state, representative of the aggregates. In all the experiments, the transition time measured by the scores of the first PC was assigned as the response value (see Table 1).

In an initial stage, the aggregation times (t_m), obtained from the scores plot of the data of all the first eight experiments (Table 1), were linearly correlated with the values of the biochemical variables according to the Eq. (1). The coefficients of the model are shown in Table 2. The difference between the aggregation times predicted by the model at the central points of the experimental domain ($\bar{t}_{m,c,predicted}$= 32.63 and 39.88 for monomeric and hexameric insulin, respectively) and the average of the responses of the experiments in these conditions ($\bar{t}_{m,c,measured}$= 24.4 and 25.8 for monomeric and hexameric insulin, respectively) is higher. This observation was statistically assessed by an F-test, comparing $SS_{pure\ quadratic}$ and $SS_{residual}$ (see Eq. (2)). The F_{cal} was 320.3 and 329.5 for monomeric and hexameric insulin and the $F_{tab\ (1,2,0.05,2tails)}$= 38.5. Therefore, $SS_{pure\ quadratic}$ and $SS_{residual}$ were not comparable, which means that a quadratic curvature is present and the model has to be represented by a second-order equation.

Table 2. Coefficients of the first-order model for both types of insulin.

	Monomeric	***Hexameric***
b_0: *average*	32.63	39.88
b_1: *T*	-23.25	-12.75
b_2: *pH*	-10.75	21.75
b_3 :*I*:	-14.25	14.75
b_{12}: *T x pH*	5.75	-5.75
b_{13}: *T x I*	7.25	-1.75
b_{23}: *pH x I*	11.75	-0.25

A second-order model was established by using the t_m of the eighteen complete experiments shown in Table 1, according to Eq. (3):

Monomeric: $t_m = 25.75 - 26.40\ x_1 - 6.80\ x_2 - 12.20\ x_3 + 12.40\ x_1^2 - 5.75\ x_1 x_2 - 7.25\ x_1 x_3 - 3.65\ x_2^2 - 11.75\ x_2 x_3 + 5.39\ x_3^2$

Hexameric: $t_m = 25.92 - 14.60\ x_1 + 18.80\ x_2 + 11.80\ x_3 + 40.31\ x_1^2 - 5.75\ x_1 x_2 - 1.75\ x_1 x_3 - 8.71\ x_2^2 - 0.25\ x_2 x_3 - 3.71\ x_3^2$

where t_m is the aggregation time and x_1 x_2 and x_3 are the codified values for temperature, pH and ionic strength, respectively.

In both cases, we can see that temperature (x_1) is the most influential variable, because its coefficient is the highest. The second most influential variable depends on the type of insulin. In monomeric insulin it is ionic strength, and, although pH is less important, its interaction with ionic strength has the same magnitude as the value of the coefficient of T. On the other hand, pH is the most influential variable in hexameric insulin, but ionic strength is also influential as its quadratic term is the same size as pH. Now, the aggregation time predicted by the model at the central point of the experimental domain ($\bar{t}_{m,c,predicted}$= 24.25 and 22.36 for monomeric and hexameric insulin, respectively) falls into the interval of the experimental values ($\bar{t}_{m,c,measured}$= 25.75 ± 3.10 and 25.92 ± 4.96 for monomeric and hexameric insulin, respectively) which indicates that the aggregation time of insulin in the range of these biochemical variables can be modelled by a second-order model.

Table 3. Experimental conditions measured and predicted aggregation time of the external samples for both types of insulin. Regression parameters for both types of insulin.

	Monomeric					***Hexameric***				
	T	pH	I	$t_{measured}$	$t_{predicted}$	T	pH	I	$t_{measured}$	$t_{predicted}$
Sample 1	37	4	0.40	67	65	37	4	0.40	30	30
Sample 2	37	5	0.50	31	28	37	5	0.50	13	15
Sample 3	38	6	0.35	43	43	33	5	0.40	46	49
Sample 4	40	4	0.55	14	16	40	4	0.55	41	44
Sample 5	42	6	0.60	38	36	36	7	0.60	35	38
Sample 6	43	5	0.30	12	13	43	5	0.30	38	43
Sample 7	44	3	0.45	33	36	44	3	0.45	74	72
Sample 8	37	3	0.60	23	23	37	3	0.60	42	46

Slope (b_1)	1.02	0.99
Intercept (b_o)	-1.38	-4.28
R^2	0.98	0.99
F_{cal}	1.37	3.18
$F_{tab\ (1,\ 2,\ \alpha=0.05)}$ = 5.14		

The suitability of the model proposed to the experimental data can be evaluated by analysing the trend of the residuals. In neither case do the residuals present any obvious pattern or trend (results not shown).

To assess the absence of bias in the models eight new experiments were performed within the experimental domain (for the conditions, see Table 3). The aggregation times obtained from their respective first scores plots ($t_{m,meas}$) were regressed against the predicted values calculated with the model for

monomeric and hexameric insulin. The regression parameters calculated using the least squares method, are also shown in Table 3. The joint confidence interval test of the slope and intercept proved that the method is free from bias at a 99.95% level of significance. In both cases, F_{cal} was lower than F_{tab}. The standard deviation (RMSEP) of the differences between the measured and the predicted results (see Table 3) were 1.96 for the monomeric and 2.66 for the hexameric insulin, respectively. Both values are comparable to the standard deviations representative of the experimental error (±3.10 and ±4.96 for monomeric and hexameric insulin, respectively), so we can assume that the accuracy of the results is acceptable.

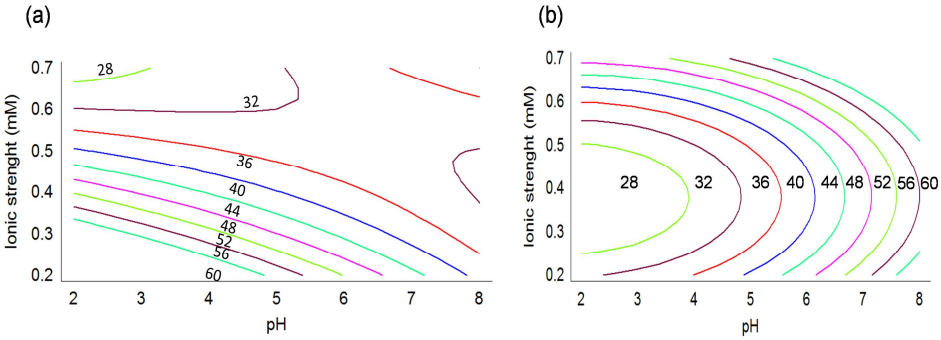


Figure 2. Contour plot at T=36ºC for: (a) monomeric and (b) hexameric insulin.

The contour map of aggregation time at human body temperature (T=36ºC) is shown in Figures 2a and 2b, for monomeric and hexameric insulin respectively. The contour map for monomeric insulin (Fig. 2a) indicates that the optimal values correspond to low values of ionic strength (I=0.20-0.33 mM) and acid pH (from 2-4.5). Figure 3a and 3b shows the optical images of aggregates, which can be considered to be representative of conditions in which aggregation time is long (Fig. 3a) and short (Fig. 3b). Overall, it can be seen that at acid pH (Fig. 3a) the aggregates are more defined, compatible with a slower formation of aggregates.

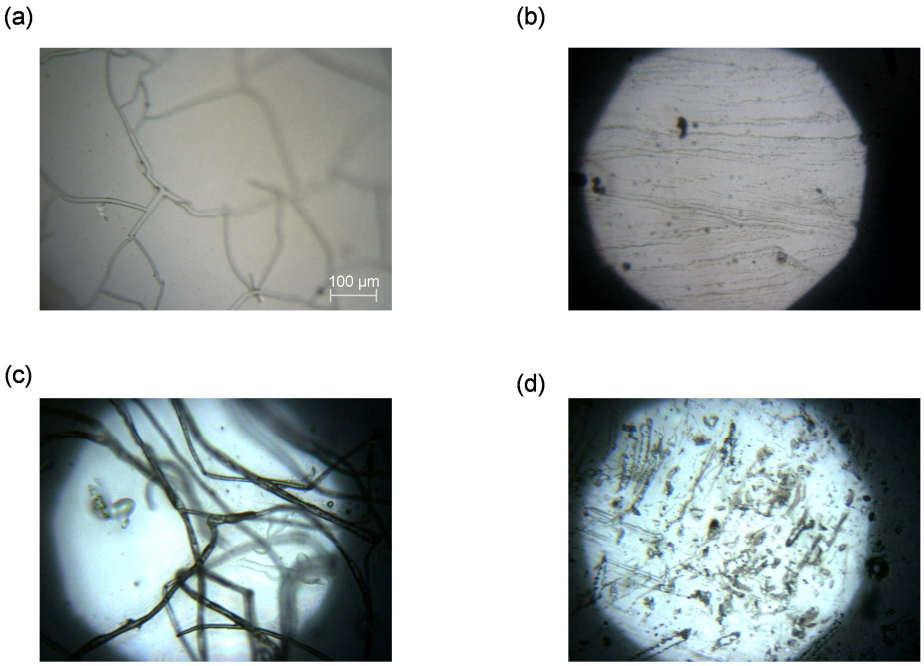


Figure 3. Optical images at the end of the process of two different conditions for monomeric ((a) and (b)) and hexameric insulin ((c) and (d)).

The contour map for hexameric insulin (Fig. 2b) shows that regardless of the ionic strength, aggregation times can be longer at high pH (6.5-8), in agreement with the literature [7,11]. Again, from the optical images representative of two different conditions (high, Fig. 3c and low aggregation time, Fig. 3d), we can see that the aggregates generated from the experimental conditions that delay aggregation, are well-defined and broader.

CONCLUSIONS

The strategy used in the present work is an easy way of finding predictive models among the values of the biochemical variables that affect the stability of insulin and its aggregation. In the experimental ranges considered (pH = 3-7,

T = 35-45ºC and I = 0.3-0.6mM), the relation between temperature (T), pH and ionic strength (I) and the aggregation time for monomeric or hexameric insulin solution follows a second-order model.

Temperature was the most influential variable in the aggregation process, independently of the type of insulin studied.

The second most influential variable depends on the type of insulin. In monomeric insulin it is ionic strength and pH is the most influential variable in hexameric insulin. The aggregation time in the monomeric insulin is influenced by the interaction between pH and ionic strength. For the hexameric insulin, the most important interaction is T and pH.

At human body temperature (T=36ºC), low values of ionic strength (I=0.2-0.33 mM) and acid pH (from 3-3.8) delay the formation of aggregates in monomeric insulin solution. This occurs at high pH (6.5-8) regardless of the ionic strength of the hexameric insulin solution.

A well-defined morphology of the aggregates was observed in the experimental conditions in which the aggregation time was the highest.

ACKNOWLEDGMENTS. We would like to thank Universitat Rovira i Virgili for providing Idoia Martí-Aluja with a doctoral fellowship.

REFERENCES

[1] Y. Hong, L. Meng, S. Chen, C. W. T. Leung, L.-T. Da, M. Faisal, D-A. Silva, J. Liu, J. W. Y. Lam, X. Huang, B. Z. Tang, *J. Am. Chem. Soc.* **134** (2012) 1680 – 1689.

[2] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky, A. L. Fink, *Biochemistry* **40** (2001) 6036 – 6046.

[3] T. J. Gibson, R. M. Murphy, *Protein Sci*. **15** (2006) 1133 **–** 1141.

[4] D-P. Hong, A. Ahmad, A. L. Fink, *Biochemistry* **45** (2006) 9342 – 9353.

[5] Y. Okada, K. Yokono, A. Katsuta, M. Yoshida, S. Morita, H. Irino, T. Goto, S. Baba, R. A. Roth, K. Shii, *Anal. Biochem*. **257** (1998) 134 – 138.

[6] J. L. Jimenez, E. N. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **99** (2002) 9196 – 9201.

[7] J. Brange, L. Andersen, E. D. Laursen, G. Meyn, E. Rasmussen, *J. Pharm. Sci*. **86** (1997) 517 – 525.

[8] J. Brange, B. Skelbaek-Pedersen, L. Langkjaer, U. Damgaard, H. Ege, S. Havelund, L. G. Heding, K. H. Jorgensen, J. Lykkeberg, J. Markussen, M. Pingel, E. Rasmussen, *Galenics of Insulin. The physicochemical and pharmaceutical aspects of insulin and insulin preparations,* Springer-Verlag, Berlin, 1987.

[9] S. Nakazawa N. Hashii, A. Harazono, N. Kawasak, *Anal. Biochem.* **420** (2012) 61 – 67.

[10] V. Toledo-Rubio, E. Vazquez, G. Platas, J. Domingo-Espín, U. Unzueta, E. Steinkamp, E. García-Fruitós, N. Ferrer-Miralles, A. Villaverde, *J. Biomol. Screen.* **15** (2010) 453 – 457.

[11] Q. Hua, M. A. Weiss, *J. Biol. Chem.* **279** (2004) 21449 – 21460.

[12] C. M. Dobson, *Nature* **18** (2003) 884 – 890.

[13] G. Zandomeneghi, M. R. H. Krebs, M. Mccammon, M. Fändrich, *Prot. Sci.* **13** (2004) 3314 – 3321.

[14] M. I. Smith, J. S. Sharp. Brange, L. Andersen, E. D. Laursen, G. Meyn, E. Rasmussen, *J. Pharm. Sci.* **86** (1997) 517 – 525.

[15] J. L. Whittingham, D. J. Scott, K. Chance, A. Wilson, J. Finch, J. Brange, G. Dodson, *J. Mol. Biol.* **318** (2002) 479 – 490.

[16] T. D. Gerlough, R. W. Bates, *J. Pharmacol. Exp. Ther.* **45** (1932) 19 – 51.

[17] V. du Vigneaud, R. H. Sifferd, R. R. Seaclock*, J. Biol. Chem.* **102** (1933) 521 – 533.

[18] M. Manno, E.F. Craparo, A. Podestà, D. Bulone, R. Carrotta, V. Martorana, G. Tiana, P.L. San Biagio, *J. Mol. Biol*. **366** (2007) 258 – 74.

[19] H. Levine, *Protein Sci.* **2** (1993) 404 – 410.

[20] C. Rodríguez-Rodríguez, A. Rimola, L. Rodríguez-Santiago, P. Ugliengo, A. Álvarez-Larena, H. Gutiérrez-de-Terán, M. Sodupe, P. González-Duarte, *Chem. Commun.* **46** (2010) 1156 – 1158.

[21] P. P. Bose, U. Chatterjee, L. Xie, J. Johansson, E. Göthelid, P. I Arvidsson, *ACS Chem. Neurosci.* **1** (2010) 315 – 324.

[22] Y. Xu, Y. Yan, D. Seeman, L. Sun, P. L. Dubin, *Langmuir* **28** (2012) 579 – 586.

[23] M. Muzaffar, A. Ahmad, *PLOS One* **6** (2011) 1 – 10.

[24] K. Huus, S. Havelund, H. B. Olsen HB, M. van de Weert, S. Frokjaer, *Biochemistry* **44** (2005) 11171 – 11177.

[25] G. T. Webster, J. Dusting, S. Balabani, E. W. Blanch, *J. Phys. Chem B* **115** (2011) 2617 – 2626.

[26] C. Ortiz, D. Zhang, A. E. Ribbe, Y. Xie, D. Ben-Amotz, *Biophys. Chem.* **128** (2007) 150 – 155.

[27] K. Huang, J. Dong, N. B. Phillips, P. R. Carey, M. A. Weiss, *J. Biol. Chem.* **280** (2005) 42345 – 42355.

[28] N. Rubin, E. Perugia, M. Goldschmidt, M. Fridkin, L. J. Addadi*, J. Am. Chem. Soc.* **130** (2008) 4602 – 4603.

[29] E. Kachooei, A. A. Moosavi-Movahedi, F. Khodagholi, H. Ramshini, F. Shaerzadeh, N. Sheibaniet, *PLoS ONE* **7** (2012) e41344.

[30] W. D. Lougheed, H. Woulfe-Flanagan, J. R. Clement, A. M. Albisser, *Diabetologia* **19** (1980) 1 – 9.

[31] R. Jansen, W. Dzwolak, R. Winter, *Biophys. J.* **88** (2005) 1344 – 1353.

[32] D. F. Waugh, D. F. Wilhelmson, S. L. Commerford, M. L. Sackler, *J. Am. Chem. Soc.* **75** (1953) 2592 – 2600.

[33] W. Wang, S. Sandeep, D. Teagarden, *Int. J. Pharm*. **390** (2010) 89 – 99.

[34] J. Winter, L. Hauke, R. Rudolph, *Anal. Biochem*. **310** (2002) 148 – 155.

[35] R. S. Lord, F. Gubensek, J. A. Rupley, *Biochemistry* **12** (1973) 4385 – 4392.

[36] W. Kadima, L. Øgendal, R. Bauer, N. Kaarsholm, K. Brodersen, J. F. Hansen, P. Porting, *Biopolymers* **33** (1993) 1643 – 1657.

[37] I. Martí-Aluja, M. S. Larrechi, *Chemom. Intell. Lab. Syst.* **127** (2013) 49 – 54.

[38] D. C. Montgomery, *Design and analysis of experiments,* Wiley, Hoboken, 2001.

[39] I. Martí-Aluja, J. Ferré, I. Ruisánchez, M. S. Larrechi, *Chemom. Intell. Lab. Syst.* **118** (2012) 180 – 186.

[40] I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi, *Anal. Chim. Acta* **760** (2013) 16 – 24.

[41] The Mathworks, MATLAB Version 7.0, Natick, MA, 2004.

[42] Statgraphics <http://www.statgraphics.com/>

[43] J. Fleming, B. Neidhart, H. Albus, W. Wegscheider, *Accredit. Qual. Assur*. **3** (1996) 135 – 138.

[44] V. Dzwolak, V. Smirnovas, R. Ansen, R. Winter, *Prot. Sci.* **13** (2004) 1927 – 1932.

[45] L. F. Pease, M. Sorci, S. Guha, D-H. Tsai, M. R. Zachariah, M. J. Tarlov, G. Belfort, *Biophys. J.* **99** (2010) 3979 – 3985.

[46] R. W. Sarver Jr., W. C. Krueger, *Anal. Biochem.* **194** (1991) 89 – 100.

[47] A. Elliot, E. J. Ambrose, *Nature* **165** (1950) 921 – 922.

[48] S. Krimm, J. Bandekar, *Adv. Prot. Chem*. **38** (1986) 181 – 364.

[49] M. Bouchard, J. Zurdo, E. J. Nettleton, C. M. Dobson, C. V. Robinson, *Prot. Sci*. **9** (2000) 1960 – 1967.

5. 1. General conclusions

Analytical methodologies based on the chemometric analysis of infrared spectra recorded during insulin aggregation are useful to analyse the evolution of this process. The conclusions drawn throughout this chapter can be divided into these points:

- Infrared spectroscopy gives information about the structural changes during the aggregation process.

- Principal component analysis (PCA) of the infrared spectra collected during the process has provided the time, detectable by infrared spectroscopy, at which insulin starts to aggregate.

- Multivariate curve resolution alternating least squares (MCR-ALS) of the infrared spectra have allowed obtaining information about the mechanism that follows insulin in presence of different external factors, such as antiretroviral drugs or different medium conditions.

- The conjunction of experimental design techniques and chemometric treatment has proved a useful tool for the analysis of processes involving insulin depending on the medium conditions.

- Response surface strategy allows establishing a model to evaluate the stability of insulin in function of the medium conditions.

5. 2. References

[1] R. E. Humbel, *Proc. Natl. Acad. Sci. U S A.* **53** (1965) 853 – 859.

[2] P. T. Grant, K. B. Reid, *Biochem. J.* **109** (1968) 31P – 32P.

[3] J. B. Field, *Acta Diabetol. Lat.* **5** (1968) 124 – 130.

[4] Y. Hong, L. Meng, S. Chen, C. W. T. Leung, L.-T. Da, M. Faisal, D-A. Silva, J. Liu, J. W. Y. Lam, X. Huang, B. Z. Tang, *J. Am. Chem. Soc.* **134** (2012) 1680 – 1689.

[5] M. Manno, E.F. Craparo, A. Podestà, D. Bulone, R. Carrotta, V. Martorana, G. Tiana, P.L. San Biagio, *J. Mol. Biol*. **366** (2007) 258 – 274.

[6] H. Levine, *Protein Sci*. **2** (1993) 404 – 410.

[7] C. Rodríguez-Rodríguez, A. Rimola, L. Rodríguez-Santiago, P. Ugliengo, A. Álvarez-Larena, H. Gutiérrez-de-Terán, M. Sodupe, P. González-Duarte, *Chem. Commun.* 46 (2010) 1156 – 1158.

[8] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky, A. L. Fink, *Biochemistry* **40** (2001) 6036 – 6046.

[9] P. P. Bose, U. Chatterjee, L. Xie, J. Johansson, E. Göthelid, P. I Arvidsson, *ACS Chem. Neurosci*. **1** (2010) 315 – 324.

[10] Y. Xu, Y. Yan, D. Seeman, L. Sun, P. L. Dubin, *Langmuir* **28** (2012) 579- 586.

[11] D. N. Brems , P. L. Brown , L. A. Heckenlaible , B. H. Frank, *Biochemistry* **29** (1990) 9289 -9293.

[12] M. Muzaffar, A. Ahmad, *PLOS One* **6** (2011) 1- 10.

[13] K. Huus, S. Havelund, H. B. Olsen HB, M. van de Weert, S. Frokjaer, *Biochemistry* **44** (2005) 11171 – 11177.

[14] G. T. Webster, J. Dusting, S. Balabani, E. W. Blanch, *J. Phys. Chem B* **115** (2011) 2617 – 2626.

[15] C. Ortiz, D. Zhang, A. E. Ribbe, Y. Xie, D. Ben-Amotz, *Biophys. Chem.* **128** (2007) 150 – 155.

[16] C. Ortiz, D. Zhang, Y. Xie, V. Jo Davisson, D. Ben-Amotz, *Anal. Biochem*. **332** (2004) 245 – 252.

[17] D. Kurouski, T. Deckert-Gaudig, V. Deckert, I. K. Lednev, *J. Am. Chem. Soc.* **134** (2012) 13323 – 13329

[18] N. Rubin, E. Perugia, M. Goldschmidt, M. Fridkin, L. J. Addadi*, J. Am. Chem. Soc.* **130** (2008) 4602 – 4603.

[19] E. Kachooei, A. A. Moosavi-Movahedi, F. Khodagholi, H. Ramshini, F. Shaerzadeh F, N. Sheibaniet, *PLoS ONE* **7** (2012) e41344.

[20] D-P. Hong, A. Ahmad, A. L. Fink, *Biochemistry* **45** (2006) 9342 – 9353.

[21] J. Koolman, K-H. Röehm, *Color atlas of biochemistry*, Georg Thieme Verlag, Rüdigerstrasse, 2005.

[22] U. Derewenda, Z. Derewenda, G. G. Dodson, R. E. Hubbard, F. Korber, *Brit. Med. Bull.* **45** (1989) 4 – 18.

[23] M. Federwisch, M. L. Dieken, P. De Meyts, *Insulin & related protein - Structure to function and pharmacology,* Kluwer Academic Publishers, Dordrecht, 2002.

[24] S. Nakazawa N. Hashii, A. Harazono, N. Kawasak, *Anal. Biochem.* **420** (2012) 61-67.

[25] J. D. Schmit, K. Ghosh, K. Dill, *Biophys. J.* **100** (2011) 450 – 458.

[26] A. B. Chausmer, *J. Am. Coll. Nutr.* **17** (1998) 109 – 115.

[27] M. Manno, E. F. Craparo, V. Martorana, D. Bulone, P. L. San Biagio, *Biophys. J*. **90** (2006) 4585 – 1591.

[28] J. Brange, L. Andersen, E. D. Laursen, G. Meyb, E. Rasmussen*, J. Pharm. Sci.* **86** (1997) 517 – 525.

[29] X. Chang, A. M. M. Jørgensen, P. Bardrum, J. J. Led, *Biochemistry* **36** (1997) 9409 – 9422.

[30] E. J. Dodson, G. G. Dodson, D. C. Hodgkin, C. D. Reynolds, *Can. J. Biochem. Cell. B.* **57** (1979) 469 – 479.

[31] E. J. Nettleton, P. Tito, M. Sunde, M. Bouchard, C. M. Dobson, C.V. Robinson, *Biophys. J.* **79** (2000) 1053 – 1065.

[32] P. D. Jeffrey, B. K. Milthorpe, L. W. Nichol, *Biochemistry* **15** (1976) 4660 – 4665.

[33] A. K. Attri, C. Fernández, A. P. Minton, *Biophys. Chem.* **148** (2010) 28 – 33.

[34] M. E. M. Cromwell, E. Hilario, F. Jacobson, *AAPS J.* **8** (2006) E572 – E579.

[35] E. Y. Chi, S. Krishnan, T. W. Randolph, J. F. Carpenter, *Pharm. Res*. **20** (2003) 1325 – 1336.

[36] A. K. Attri, C. Fernádez, A. P. Minton, *Biophys. Chem*. **148** (2010) 23 – 27.

[37] W. Dzwolak, A. Loksztejn, V. Smirnovas, *Biochemistry* **45** (2006) 8143 – 8151.

[38] T. Arakawa, J. S. Philo, D. Ejima, K. Tsumoto, F. Arisaka, *Bioprocess Int*. **4** (2006) 42 – 43.

[39] M. I. Ivanova, S. A. Sievers, M. R. Sawaya, J. S. Wall, D. Eisenberg*, Proc. Natl. Acad. Sci. USA* **106** (2009) 18990 – 18995.

[40] V. Toledo-Rubio, E. Vazquez, G. Platas, J. Domingo-Espín, U. Unzueta, E. Steinkamp, E. García-Fruitós, N. Ferrer-Miralles, A. Villaverde, *J. Biomol. Screen*. **15** (2010) 453 – 457.

[41] Q. Hua, M. A. Weiss, *J. Biol. Chem*. **279** (2004) 21449 – 21460.

[42] G. Zandomeneghi, M. R. H. Krebs, M. Mccammon, M. Fändrich, *Prot. Sci.* **13** (2004) 3314 – 3321.

[43] V. N. Uversky, A. L. Fink, *Biochim. Biophys. Acta* **1698** (2004) 131 – 153.

[44] C. M. Dobson, *Nature* **18** (2003) 884 – 890.

[45] A. Ahmad, V. N. Uversky, D. Hong, A. L. Fink*, J. Biol. Chem*. **280** (2005) 42669 – 42675.

[46] K. Giger, R. P. Vanam, E. Seyrek, P. L. Dublin, *Biomacromolecules* **9** (2008) 2338 – 2344.

[47] J. Brange, L. Langkjoer, *Pharm. Biotechnol*. **5** (1993) 315 – 350.

[48] V. Smirnovas, R. Winter, *Biophys. J.* **94** (2008) 3241 – 3246.

[49] V. Montessori, N. Press, M. Harris, L. Akagi, J. S. G. Montaner, *Can. Med. Assoc. J.* **179** (2004) 229 – 238.

[50] C. Grunfeld, *Clin. Infect. Dis.* **47** (2008) 575 – 576.

[51] W. Ismail, J. A. King, T. S. Pillay, *JEMDSA* **14** (2009) 129 – 132.

[52] M. Aboud, A. Elgalib, R. Kulasegaram, B. Peters, *Int. J. Clin. Pract*. **61** (2007) 463 – 472.

[53] G. M. Reaven, *Diabetes* **37** (1988) 1595 – 1607.

[54] G. I. Shulman, *J. Clin. Invest*. **106** (2000) 171 **–** 176.

[55] G. Dodson, D. Steiner*, Curr. Opin. Struct. Biol*. **8** (1998) 189 – 194.

[56] J. G. Menting, J. Whittaker, M. B. Margetts, L. J. Whittaker, GK-W. Kong, B. J. Smith, C. J. Watson, L. Žáková, E. Kletvíková, J. J. Chan, S. J. Steiner, D. F. Dodson, G. G. Brzozowski, A. M. Weiss, C. W. Ward, M. C. Lawrence, *Nature* **493** (2013) 241 – 245.

[57] Q. Hua, *Protein Cell* **1** (2010) 537 – 551.

[58] M. Kanzaki, *Endocr. J.* **53** (2006) 267 – 293.

[59] F. Kanai, K. Ito, M. Todaka, H. Hayashi, S. Kamohara, K. Ishii, T. Okada, O. Hazeki, M. Ui, Y. Ebina*, Biochem. Bioph. Res. Co.* **195** (1993) 762 – 768.

[60] C. Lagathu, M. Kim, M. Maachi, C. Vigouroux, P. Cervera, J. Capeau, M. Caron, J-P. Bastard, *Biochimie* **87** (2005) 65 – 71.

[61] M. Fasshauer, R. Paschke, *Diabetologia* **46** (2003) 1594 – 1603.

[62] P. A. Cane, *J. Antimicrob. Chemother*. **64** (2009) 37 – 40.

[63] A. Checa, R. Oliver, S. Hernández-Cassou, J. Saurina, *Anal. Chim. Acta* **647** (2009) 1 – 13.

[64] M. Dybul, A. S. Fauci, J. G. Bartlett, J. E. Kaplan, A. K. Pau, *Ann. Intern. Med.* **137** (2002) 381 – 433.

[65] V. Jain, S. G. Deeks, *Curr. HIV/AIDS Rep.* **7** (2010) 60 – 68.

[66] A. Thorner, E. Rosenberg, *Drugs* **63** (2003) 1325 – 1337.

[67] M. A. Fischl, *AIDS* **13** (1999) S49 – S59.

[68] J. S. Philo, T. Arakawa, *Curr. Pharm. Biotech*. **10** (2009) 348 – 351.

[69] R. Jansen, W. Dzwolak, R. Winter, *Biophys J.* **88** (2005) 1344 – 1353.

[70] J. L. Whittingham, D. J. Scott, K. Chance, A. Wilson, J. Finch, J. Brange, G. Dodson, *Mol. Biol.* **318** (2002) 479 – 490.

[71] J. Winter, L. Hauke, R. Rudolph, *Anal. Biochem*. **310** (2002) 148 – 155.

[72] A. Ahmad, I.S. Millett, S. Doniach, V. N. Uversky, A. L. Fink, *Biochemistry* **42** (2003) 11404 – 11416.

[73] D. F. Waugh, D. F. Wilhelmson, S. L. Commerford, M. L. Sackler, *J. Am. Chem. Soc*. **75** (1953) 2592 – 2600.

[74] J. Haas, E. Vöhringer-Martinez, A. Andreas Bögehold, D. Matthes, U. Hensen, A. Pelah, B. Abel, H. Grubmüller, *ChemBioChem*. **10** (2009) 1816 – 1822.

[75] C. L. Heldt, M. Sorci, D. Posada, A. Hirsa, G. Belfort, *Biotechnol. Bioeng*. **108** (2011) 237 – 241.

[76] T. P. J. Knowles, W. Shu, G. L. Devlin, S. Meehan, S. Auer, C. M. Dobson, M. E. Welland, *Proc. Natl. Acad. Sci. USA* **104** (2007) 10016 – 10021.

Chapter 6

General conclusions

6. 1. Conclusions

This chapter summarises the main conclusions of the studies presented in this doctoral thesis. More specific conclusions can be found at the end of each research paper.

Infrared spectroscopic techniques combined with chemometric treatment is a valuable alternative to more complex and expensive analytical techniques, as they have proven to be useful for analysing a process of polymerisation and aggregation of insulin.

- A polymerisation process can be monitored *in situ* by near-infrared chemical imaging (NIR-CI), which detects the chemical changes of this process in real time.

- Chemometric analysis of the NIR images recorded during the polymerisation process gives useful information not only about the chemical changes involved in the process but also about the spatial distribution of the polymer chain during the polymerisation process.

- Structural changes during the formation of insulin aggregates can be monitored *in situ* by infrared spectroscopy.

- Analytical methods based on principal component analysis (PCA) of the infrared spectra collected during insulin aggregation made it possible to:

    - Show the dimer dissociation from the monomer forms.

- Quantify the effect of three antiretroviral drugs on the kinetics of insulin aggregation

- Analytical methods based on the multivariate curve resolution alternating least squares (MCR-ALS) of the infrared spectra recorded during insulin aggregation in the presence of antiretroviral drugs (zidovudine, ritonavir and efavirenz) have been shown to be a powerful tool to :

  - Obtain quantitative evidence of the effect of antiretroviral drugs on aggregation over time.

  - Extract additional information about the pathway along which the process occurs.

- Experimental design techniques reduce the number of experiments required to simultaneously analyse the influence of the medium conditions on insulin aggregation. These methods in conjunction with the chemometric treatment of the infrared spectra collected made it possible to:

  - Measure the influence of some biochemical variables (pH, temperature and ionic strength) on the aggregation time.

  - Quantify the interaction between the biochemical variables.

  - Model insulin aggregation time in function of temperature, pH and ionic strength of two types of insulin (monomeric and hexameric).

- Multivariate curve resolution-alternating least squares analysis of the infrared spectra collected during insulin aggregation in different medium conditions has revealed that the medium conditions can affect the mechanisms that occur in insulin.

Appendix

APPENDIX A: List of abbreviations

Abs	Absorbance
AIDS	Acquired Immune Deficiency Syndrome
ALS	Alternating Least Squares
ANOVA	Analysis of Variance
ATR	Attenuated Total Reflection
AZT	3'-azido-3'-deoxythymidine
CD	Circular Dicroism
Corrcoef	Correlation coefficient
DDM	Diaminodiphenyl methane
EFA	Evolving Factor Analysis
ENO	Enone-containing triglyceride
FPA	Focal Plane Array
FSIW-EFA	Fixed Size Image Window-Evolving Factor Analysis
FSMW-EFA	Fixed Size Moving Window-Evolving Factor Analysis
FTIR-ATR	Fourier Transform infrared spectroscopy-attenuated total reflection
HAART	Highly Active Antiretroviral Therapy
LOF	Lack of Fit
MCR-BANDS	Multivariate Curve Resolution-BANDS
MCR-ALS	Multivariate Curve Resolution-Alternating Least Squares
NIR-CI	Near-Infrared Chemical Imaging
NMR	Nuclear Magnetic Resonance
PCA	Principal Components Analysis
PC	Principal Components
SEM	Scanning Electron Microscopy
SIMPLISMA	SIMPLe-to-use Interactive Self-modeling Mixture Analysis
SNV	Standard Normal Variate
SVD	Singular Value Decomposition
SV	Singular Value
UV-Vis	Ultraviolet Visible

APPENDIX B: List of papers by the author presented in this thesis

1. I. Martí-Aluja, I. Ruisánchez, V. Cádiz, S. Maspoch and M. S. Larrechi. Aza-Michael reaction with enone-modified vegetable oils: evidence of the keto-enolic equilibrium by NIR chemical imaging and evolving factor analysis. Analytical and Bioanalytical Chemistry, 399 (2010) 1975-1982.

2. I. Martí-Aluja, J. Ferré, I. Ruisánchez and M. S. Larrechi. *Chemometrics analysis of insulin aggregation induced by an antiretroviral drug (AZT).* Chemometrics and Intelligent Laboratory Systems, 118 (2012) 180-186.

3. I. Martí-Aluja, I. Ruisánchez and M. S. Larrechi. Quantitative analysis of the effect of zidovudine, efavirenz and ritonavir on insulin aggregation by multivariate curve resolution alternating least squares. Analytica Chimica Acta, 760 (2012) 16-24.

4. I. Martí-Aluja and M. S. Larrechi. *MCR-ALS analysis of insulin aggregation / association processes. Influence of biochemical variables*. Chemometrics and Intelligent Laboratory Systems, 127 (2013) 49-54.

5. I. Martí-Aluja and M. S. Larrechi. Modelling insulin aggregation time using response surface methodology. Submitted to Chemometrics and Intelligent Laboratory Systems.

List of papers by the author not presented in this thesis

I. Martí-Aluja, S. Maspoch and M. S. Larrechi. Comparison of two near-infrared camera optical devices in chocolate candies quality control using chemical images. Paper in preparation.

APPENDIX C: Contributions to international meetings

- **Oral communications**

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi.

Zidovudine and insulin interaction study in human serum by infrared spectroscopy and chemometrics analysis. **XIII Conference on Instrumental Analysis** - Barcelona (Spain, 2011)

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi.

Extracting information by the chemometrics analysis of the AZT interaction in human serum**. IV Chemometrics workshop** - Burgos (Spain, 2011)

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi.

Studying a copolymerisation reaction from renewable sources by means of multivariate analysis of chemical images. **VI Meeting of Young Researchers from the Catalan Countries** - Valencia (Spain, 2010).

- **Poster communications**

I. Martí-Aluja, M. S. Larrechi. Evidencing the aggregation/association processes of insulin by means of MCR-ALS analysis. **VIII Colloquium Chemiometricum Mediterraneum** - Bevagna (Italy, 2013)

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi. Analysing the effect of several antiretroviral drugs on the aggregation process by MCR-ALS of infrared data. **XII Chemometrics in Analytical Chemistry** - Budapest (Hungary, 2012)

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi. Effect of the medium conditions in the insulin aggregation pathway by experimental design and MCR-ALS of infrared spectra. **XXIII National Meeting of Spectroscopy - VII Iberian Congress of Spectroscopy** - Córdoba (Spain, 2012)

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi. Aza-Michael reaction with enone-modified vegetable oils. Evidence of the keto-enolic equilibrium by NIR chemical imaging and evolving factor analysis**. VII Colloquium Chemiometricum Mediterraneum** - Granada (Spain, 2010)

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi. Study of a keto-enolic equilibrium influence in a copolymerisation reaction by near-infrared chemical imaging**. XXII National Meeting Spectroscopy - VIII Iberian Congress of Spectroscopy** - Porto (Portugal, 2010)

I. Martí-Aluja, I. Ruisánchez, M. S. Larrechi. Exploring copolymerisation reactions by near-infrared chemical. **I International Workshop on Multivariate Image Analysis** - Valencia (Spain, 2009)

UNIVERSITAT ROVIRA I VIRGILI
ANALYSIS OF POLYMERISATION / AGGREGATION PROCESSES BY NIR CHEMICAL IMAGING AND FTIR SPECTROSCOPY
Idoia Martí Aluja
Dipòsit Legal: T. 1427-2013

Tot està per fer, i tot és possible

Miquel Martí i Pol