



Universitat Autònoma de Barcelona

ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  http://cat.creativecommons.org/?page_id=184

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <http://es.creativecommons.org/blog/licencias/>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>



**Universitat Autònoma
de Barcelona**

**Implementación de técnicas analíticas rápidas para la
fabricación farmacéutica avanzada**

Judit Brassier Font

Tesis Doctoral

Programa de Doctorat de Química

Director: Dr. Manel Alcalà Bernàrdez

Departament de Química

Facultat de Ciències

2019

Memoria presentada para aspirar al Grado de Doctor por **Judit Brassier Font**

Judit Brassier Font

Visto Bueno,

Dr. Manel Alcalà Bernàrdez

Bellaterra, 3 de Mayo de 2019

Este proyecto de investigación se ha realizado en colaboración con la empresa Laboratorios Menarini S.A. en el marco del Pla de Doctorat Industrial, convocatoria DI 2015.



Responsable y tutor del proyecto en la empresa:

Josep Maria González Bosch



Badalona, 2 de Mayo de 2019

AGRADECIMIENTOS

Esta tesis doctoral se ha realizado gracias a las siguientes ayudas institucionales y empresas:

Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) de la Secretaria d'Universitat i Recerca del Departament d'Empresa i Coneixament de la Generalitat de Catalunya

Ayuda de Doctorat Industrial del proyecto 2015DI 014 con título "*Implementació de tecnologies analítiques ràpides per a la fabricació farmacèutica avançada*".

Laboratorios Menarini S.A. Badalona, Barcelona.

Aquesta tesis va començar i ha finalitzat, en gran part, gràcies a l'aportació de molta gent que d'alguna manera o altre hi ha influït. En els pròxims paràgrafs intentaré copsar tot els agraïments.

Així que començaré per agrair de tot cor als responsables de la gènesis d'aquesta Tesis, el Prof. Marcelo Blanco per haver-me donat l'oportunitat d'aprendre i de viure aquesta experiència, i al Dr. Manel Alcalà per haver-me fet de Director de Tesis. En aquesta gènesis, no voldria oblidar-me del Xavier Arnés per la confiança que sempre va tenir en mi. També vull fer extensiu l'agraïment al Prof. Santiago Maspoch i al Prof. Jordi Coello.

Aquesta tesis s'ha dividit en dos blocs o etapes, estant un any a la UAB i els darrers 3 anys a LMSA realitzant el doctorat industrial. Així que voldria donar les gràcies al Josep Maria González per la seva feina com a tutor dins l'empresa i per tot el coneixement transmès (sobretot d'un món totalment nou per mi, la indústria farmacèutica). També vull agrair a la Pilar Rajadel per haver-me acollit al departament de Control de Qualitat. A ells dos, junt al Ramón Esquerdo, voldria agrair-vos totes les reunions que hem mantingut per tal que la feina arribés a bon port.

Una de les parts més boniques de fer aquesta tesis ha estat conèixer als meus companys, tant de la UAB com de LMSA:

- Totes les experiències que vam viure als inicis (acadèmiques i sobretot extra-curriculars) amb l'Anna Palou, la Vanessa Cárdenas, Diego Alonso Gómez y Dong Sun em van fer veure que sense cap mena de dubte no m'havia equivocat amb la decisió que vaig pendre. També vull agrair a la Glòria Clua i la Laia Garcia per tota l'ajuda rebuda! També a la Judit Puig i a la Desiree Moya. Però sobretot, dono les gràcies al doctorat per haver-me permès conèixer una persona meravellosa, Aira Yira Miró amb qui he compartit aquest viatge. Estem aquí i trobarem la vall!
- I què dir de LMSA... Aquí serà més complicat donar les gràcies de manera individual sense deixar-me a ningú, ja que molta gent hi ha format part i sempre ha estat disposada a ajudar-me. Així que començo a agrair per departaments, d'aquesta manera segur que no em deixo a ningú: Control de Qualitat, Garantia de Qualitat, Tecnologia Farmacèutica, Enginyeria i Producció. Però, com sempre, hi ha gent que han tingut un major vincle amb aquesta tesis. Així que voldria donar les gràcies particularment al meus companys de dinar durant l'últim any, Víctor (per la Nutella, TW i destrossar-me series, per aquest ordre) i Batllori (per ser l'analista 2 d'aquesta tesis i el tancador de comportes oficial). A la persona que m'ha aguantat, escoltat quasi diàriament i ajudat amb el equips i derivats, Anna Vidal. A tots els companys del torn d'esmorzar, entre ells Eva i Francesc (*chromatografista profesional* ☺). Al meu company de laboratori, Juan Carlos, gracias!. En especial també m'agradaria donar les gràcies al Javi i al Juan, els operaris que més feina el hi he donat per posar-me els bins i

sempre han estat disposats a donar-me un cop de mà, sense oblidar-me de tots els que formen part de sòlids, que en major o menor mesura també els he molestat. Finalment, donar les gràcies a la Margarita Bosch, per tot el traspàs de coneixement i a dues persones clau en aquesta tesis, Andrés i Marta, gràcies QAs!

A part de la gent que té una vinculació més directe amb el dia a dia de la tesis, també vull agrair als amigotes (Núria, Carlos, Xavi, Judit, Higo, Ire i Emilio) per estar allà! Ara sí que sí, no hi ha excusa, el següent capítol que escrigui és el llibre! En especial voldria mencionar al Carlos i Nurieta, els companys de pis que em vam donar tots el ànims per començar l'aventura, i per tots el dijous és clar! Gràcies companys! També vull agrair a una altre gran persona, Pili, gràcies per estar allà sense ser-hi i per entendre aquesta última època... Les converses eternes i el teu simple “estar” m’han donat molta seguretat per continuar, t’estimo. També vull donar les gràcies a la Clara, una part d’aquest llibre és seu.

I com no podia ser d’una altre manera, sense dubte aquesta tesis li dec i és per als meus quatre pilars. Vull donar les gràcies als meus pares, Xavier i Mercè, pels sacrificis que han fet per tal de donar-nos la millor educació possible. Sobretot a tu, mama, per ser la persona que quan sigui gran m’hi vull assemblar (encara que sigui una mica). Ets la meva heroïna particular! Gràcies al meu germà Sergi, que sempre m’ha donat un punt de vista diferent de la vida. I finalment, per en Mata, no sé com donar-te les gràcies per haver estat els braços en tots els moments de feblesa en aquest camí, per haver estat el somriure que necessitava en cada moment, per ser tu. GRÀCIES!!

Tothom ha tingut una influència en el resultat d’aquesta Tesis Doctoral. De la mateixa manera que en un algoritme de reducció de variables, en aquests paràgrafs he mencionat les variables més importants, però les dades no queden 100% explicades... Totes les experiències en la vida són la suma de infinites petites aportacions, així que per si de cas, Gràcies!

Contenido

LISTA DE ACRÓNIMOS	1
RESUMEN.....	3
OBJETIVOS.....	5
Parte I	9
INTRODUCCIÓN.....	9
1. INDUSTRIA FARMACÉUTICA AVANZADA	11
1.1. ENTORNO.....	11
1.2. FABRICACIÓN TRADICIONAL DE UN MEDICAMENTO SÓLIDO	12
1.3. ENFOQUE MEJORADO DE LA FABRICACIÓN DE MEDICAMENTOS.....	15
1.3.1. Fabricación en continuo	15
1.3.2. Calidad por diseño (QbD)	16
1.3.3. Tecnología Analítica de Proceso (PAT).....	18
1.4. BIBLIOGRAFÍA.....	20
2. HERRAMIENTAS DE OBTENCIÓN DE DATOS.....	23
2.1. ESPECTROSCOPIA NIR.....	23
2.1.1. Fundamentos de la radiación NIR	23
2.1.2. Instrumentación.....	25
2.1.3. Parámetros que afectan a la señal.....	28
2.1.4. Aplicación NIR en la industria farmacéutica.....	28
2.2. ESPECTROSCOPIA RAMAN	29
2.2.1. Fundamentos.....	29
2.2.2. Instrumentación.....	31
2.2.3. Parámetros que afectan a la señal.....	33
2.2.4. Aplicación Raman en la industria farmacéutica.....	34
2.3. ESPECTROMETRÍA DE MOVILIDAD IÓNICA	35
2.3.1. Fundamentos de IMS.....	35
2.3.2. Instrumentación IMS.....	36
2.3.3. Aplicación de IMS en la industria farmacéutica.....	38
2.4. DATOS DE PROCESO	39
2.4.1. Parámetros de calidad de materias primas.....	39
2.4.2. Parámetros de proceso.....	39
2.4.3. Atributos de calidad.....	40
2.4.4. Aplicaciones en la industria farmacéutica.....	40

2.5.	BIBLIOGRAFÍA	40
3.	TÉCNICAS DE ANÁLISIS MULTIVARIABLE DE DATOS	47
3.1.	DISEÑO DE EXPERIMENTOS	47
3.2.	PRETRATAMIENTO DE DATOS	49
3.2.1.	Pretratamientos por columnas	49
3.2.2.	Pretratamientos por filas	50
3.3.	ANÁLISIS CUALITATIVO	52
3.3.1.	Análisis de componentes principales (PCA).....	53
3.3.2.	Coeficiente de correlación	54
3.3.3.	Distancias	55
3.3.4.	Análisis lineal discriminante (LDA).....	56
3.3.5.	SIMCA.....	57
3.4.	ANÁLISIS CUANTITATIVO	57
3.4.1.	Regresión parcial por mínimos cuadrados (PLS).....	58
3.5.	DESARROLLO DE UN MÉTODO NIR Y RAMAN EN EL ENTORNO FARMACÉUTICO	59
3.5.1.	Definición del objetivo.....	60
3.5.2.	Preparación del conjunto de calibración y validación.....	60
3.5.3.	Adquisición del espectro	61
3.5.4.	Determinación de la propiedad por el método de referencia	62
3.5.5.	Pretratamiento de los datos y selección rango espectral	62
3.5.6.	Exploración de los datos y detección de <i>outliers</i>	63
3.5.7.	Selección de muestras	63
3.5.8.	Modelo de calibración y optimización.....	64
3.5.9.	Validación externa	65
3.5.10.	Uso y mantenimiento del modelo	66
3.6.	BIBLIOGRAFÍA	67
Parte II	69
	MÉTODOS DE IDENTIFICACIÓN	69
4.	IDENTIFICACIÓN DE MATERIAL DE PARTIDA.....	71
4.1.	INTRODUCCIÓN	71
4.2.	MÉTODO EXPERIMENTAL.....	73
4.2.1.	Muestras	73
4.2.2.	Adquisición de espectros	73
4.2.3.	Desarrollo de la biblioteca NIR	74
4.2.4.	Desarrollo de bibliotecas Raman	77

4.3.	RESULTADOS Y DISCUSIÓN	78
4.3.1.	Biblioteca NIR.....	78
4.3.2.	Biblioteca Raman	90
4.4.	CONCLUSIONES.....	98
4.5.	BIBLIOGRAFÍA.....	99
5.	IDENTIFICACIÓN DE PRODUCTOS GRANEL.....	101
5.1.	INTRODUCCIÓN.....	101
5.2.	MÉTODO EXPERIMENTAL	102
5.2.1.	Muestras	102
5.2.2.	Adquisición de espectros.....	102
5.2.3.	Tratamiento de datos	103
5.3.	RESULTADOS Y DISCUSIÓN	103
5.3.1.	Cualificación por modo reflectancia.....	103
5.3.2.	Cualificación por modo transmisión.....	107
5.4.	CONCLUSIONES.....	109
5.5.	BIBLIOGRAFÍA.....	110
Parte III	111
ANÁLISIS EN EL PROCESO.....		111
6.	IMPLEMENTACIÓN DE UN SISTEMA PAT EN LA INDUSTRIA FARMACÉUTICA.....	113
6.1.	INTRODUCCIÓN.....	113
6.2.	ANTECEDENTES.....	115
6.3.	MÉTODO EXPERIMENTAL	116
6.3.1.	Producto.....	116
6.3.2.	Conjunto de muestras para la calibración y validación	117
6.3.3.	Adquisición de espectros.....	119
6.3.4.	Métodos de referencia	120
6.3.5.	Software.....	121
6.3.6.	Tratamiento de los datos.....	121
6.4.	RESULTADOS Y DISCUSIÓN	121
6.4.1.	Modelos predictivos	121
6.4.2.	Implementación del sistema PAT.....	133
6.5.	CONCLUSIONES.....	137
6.6.	BIBLIOGRAFÍA.....	137
7.	ESTUDIO DE VIABILIDAD DE UN MÉTODO NIR PARA IMPLEMENTAR EN LA ETAPA DE DOSIFICACIÓN.....	139
7.1.	INTRODUCCIÓN.....	139

7.2.	MÉTODO EXPERIMENTAL.....	140
7.2.1.	Muestras.....	140
7.2.2.	Adquisición de espectros	140
7.2.3.	Tratamiento de los datos	142
7.3.	RESULTADOS Y DISCUSIÓN	142
7.3.1.	Justificación de la elección de los productos analizados.....	142
7.3.2.	Exploración de los datos	144
7.3.3.	Respuesta del NIR a variaciones de concentración de API	147
7.3.4.	Detección de comprimidos rotos.....	149
7.4.	CONCLUSIONES	150
7.5.	.BIBLIOGRAFÍA	150
Parte IV	153
	VALIDACIÓN Y VERIFICACIÓN DE LIMPIEZA.....	153
8.	IMPLEMENTACIÓN DE LÍMITES PARA MÉTODOS NO ESPECÍFICOS	155
8.1.	INTRODUCCIÓN	155
8.2.	MÉTODO EXPERIMENTAL.....	156
8.2.1.	Materiales y equipo.....	156
8.2.2.	Preparación y medida de muestras.....	157
8.2.3.	Definición de los factores, de la respuesta y niveles para el DoE.....	158
8.3.	RESULTADOS Y DISCUSIÓN	159
8.3.1.	Evaluación inicial de los agentes de limpieza.....	159
8.3.2.	Estudio de la influencia de factores en la conductividad	160
8.3.3.	Establecimiento conductividad límite para verificar la limpieza	166
8.4.	CONCLUSIONES	169
8.5.	BIBLIOGRAFÍA	170
9.	CARACTERIZACIÓN DE ALCOHOLES GRASOS LINEALES ETOXILADOS POR ESPECTROMETRÍA DE MOVILIDAD IÓNICA (IMS)	171
9.1.	INTRODUCCIÓN	171
9.2.	MÉTODO EXPERIMENTAL.....	173
9.2.1.	Muestras.....	173
9.2.2.	Espectrometría de movilidad iónica.....	173
9.3.	RESULTADOS Y DISCUSIÓN	174
9.3.1.	Fundamento teórico.....	174
9.3.2.	Caracterización de los espectros de movilidad iónica.....	175
9.3.3.	Validación del modelo establecido	180
9.4.	CONCLUSIONES	183

9.5.	BIBLIOGRAFÍA.....	183
10.	CUANTIFICACIÓN DE ALCOHOLES GRASOS LINEALES ETOXILADOS POR ESPECTROMETRÍA DE MOVILIDAD IÓNICA (IMS).....	187
10.1.	INTRODUCCIÓN.....	187
10.2.	MÉTODO EXPERIMENTAL	187
10.2.1.	Muestras	187
10.2.2.	Espectrometría de movilidad iónica	188
10.2.3.	Procesado de datos	188
10.2.4.	GC-MS	189
10.3.	RESULTADOS Y DISCUSIÓN.....	189
10.3.1.	Establecimiento de parámetros de análisis y tratamiento de datos.....	189
10.3.2.	Distribución de los oligómeros.....	193
10.3.3.	Estudio cuantitativo de los oligómeros.....	195
10.4.	CONCLUSIONES.....	197
10.5.	BIBLIOGRAFÍA.....	198
Parte V	199
	ANÁLISIS DATOS DE PROCESO	199
11.	REVISIÓN HISTÓRICA DE LOS DATOS DE PROCESO.....	201
11.1.	INTRODUCCIÓN.....	201
11.2.	MÉTODO EXPERIMENTAL	202
11.2.1.	Obtención de los datos y preparación de las matrices	202
11.2.2.	Análisis de los datos	205
11.3.	RESULTADOS Y DISCUSIÓN.....	206
11.3.1.	MSPC de los parámetros de calidad.....	206
11.3.2.	Análisis de datos de proceso.....	210
11.4.	CONCLUSIONES.....	213
11.5.	BIBLIOGRAFÍA.....	214
	CONCLUSIONES.....	217

LISTA DE ACRÓNIMOS

Acrónimo	Inglés	Castellano
AGAUR	-----	Agència de Gestió d'Ajuts Universitaris i de Recerca
ANN	<i>Artificial Neural Network</i>	Redes Neuronales
AOTF	<i>Acousto-optic tunable filter</i>	Filtros acústico-óptico
API	<i>Active Pharmaceutical Ingredient</i>	Ingrediente Activo Farmacéutico
CCD	<i>Charge Coupled Devices</i>	Dispositivos de Carga Acoplada
CE	<i>Capillary Electrophoresis</i>	Electroforesis Capilar
CM	<i>Continuous Manufacturing</i>	Fabricación en Continuo
CMA	<i>Critical Material Attributes</i>	Atributos Críticos de Material
CPP	<i>Critical Process Parameters</i>	Parámetros Críticos del Proceso
CPS	<i>Cyber-Physical Systems</i>	Sistemas Cibernéticos
CQA	<i>Critical Quality Attributes</i>	Atributo Crítico de Calidad
DeEGMD	-----	Polietilenglicol Monododecil éter con promedio de 10 grupos etoxi
DEG	<i>Diethylene Glycol</i>	Dietilenglicol
DoE	<i>Design of Experiments</i>	Diseño de Experimentos
EMA	<i>European Medicine Agency</i>	Agencia Europea del Medicamento
EP	<i>European Pharmacopoeia</i>	Farmacopea Europea
ERP	<i>Enterprise Resource Planning</i>	Sistema de Planificación de Recursos Empresariales
ESI	<i>Electrospray Ionization</i>	Ionización por Electrospray
FDA	<i>Food and Drug Administration</i>	-----
FFT	<i>Fast Fourier Transform</i>	Transformada Rápida de Fourier
FIR	<i>Far-Infrared</i>	Infrarrojo Lejano
FT	<i>Fourier Transform</i>	Transformada de Fourier
GC	<i>Gas Chromatography</i>	Cromatografía de Gases
HPLC	<i>High Performance Liquid Chromatography</i>	Cromatografía Líquida de Alta Eficacia
ICH	<i>International Council for Harmonisation</i>	Consejo Internacional para la Armonización
IMS	<i>Ion Mobility Spectrometry</i>	Espectrometría de Movilidad Iónica
IoS	<i>Internet of Services</i>	Internet de los Servicios
IoT	<i>Internet of Things</i>	Internet de las Cosas
IPC	<i>In Process Control</i>	Control en Proceso
IR	Infrared	Infrarrojo
LAE	<i>Lineal Alcohol Ethoxylate</i>	Alcohol Lineal Etoxilado
LD	Linear Discriminant Function	funciones lineales discriminantes
LD ₅₀	<i>Median Lethal Dose</i>	Dosis Letal Media
LDA	<i>Linear Discriminant Analysis</i>	Análisis Lineal Discriminante
LED	<i>Light-Emitting Diode</i>	Diodo Emisor de Luz
LIMS	<i>Laboratory Information Management System</i>	Sistema de Gestión de Información de Laboratorio
LOD	<i>Limit of Detection</i>	Límite de Detección
MACO	<i>Maximum Allowable Carryover</i>	Máximo Remanente Permitido
MALDI	<i>Matrix-assisted laser desorption/ionization</i>	Desorción/Ionización láser asistida por matriz
MBS	<i>Minimum Batch Size</i>	Tamaño de Lote Mínimo
MCC	<i>Microcrystalline Cellulose</i>	Celulosa microcristalina
MIR	<i>Mid-Infrared</i>	Infrarrojo Medio

LISTA DE ACRÓNIMOS

Acrónimo	Inglés	Castellano
MLR	<i>Multiple Linear Regression</i>	Regresión Lineal Múltiple
MPCA	<i>Multiway PCA</i>	-----
MS	<i>Mass Spectrometry</i>	Espectrometría de Masas
MSC	<i>Multiplicative Scattering Correction</i>	Corrección del Efecto Multiplicativo de Dispersión
MSPC	<i>Multiple Statistical Process Control</i>	Control Estadístico Múltiple del Proceso
NCF o GMP	<i>Good Manufacturing Practices</i>	Normas de Correcta Fabricación
NIR	<i>Near-Infrared</i>	Infrarrojo Cercano
OOS	<i>Out Of Specification</i>	Fuera de Especificaciones
OT	<i>NA</i>	Orden de Trabajo
PAT	<i>Process Analytical Technology</i>	Tecnología Analítica de Proceso
PC	<i>Personal Computer</i>	Ordenador Personal
PCA	<i>Principal Component Analysis</i>	Análisis de Componentes Principales
PCR	<i>Principle Component Regression</i>	Regresión por Componentes Principales
PDE	<i>Permitted Daily Exposure</i>	Exposición Diaria Permitida
PEG	<i>Polyethylene Glycol</i>	Polietilenglicol
PLC	<i>Programmable Logic Controller</i>	Controlador Lógico Programable
PLS	<i>Partial Least Squares Regression</i>	Regresión Parcial por Mínimos Cuadrados
PQR	<i>Product Quality Review</i>	Informe Periódico de calidad
PRM	<i>Pattern Recognition Methods</i>	Métodos de Reconocimiento de Patrones
QbD	<i>Quality by Design</i>	Calidad por Diseño
QTPP	<i>Quality Target Product Profile</i>	Perfil de la Calidad Objetivo del Producto
RMSECV	<i>Root Mean Squared Error Cross-Validation</i>	Error Cuadrático Medio de Validación Cruzada
RMSEP	<i>Root Mean Squared Error Prediction</i>	Error Cuadrático Medio de Predicción
RSE	<i>Root Standard Error</i>	Error Estándar Relativo
RSD	<i>Relative Standard Deviation</i>	Desviación Estándar Relativa
RTRT	<i>Real Time Release Testing</i>	Ensayo de Liberación en Tiempo Real
SEL	<i>Standard Error of Laboratory</i>	Error Estándar de Laboratorio
SF	<i>Security Facto</i>	Factor de seguridad
SIM	<i>Single Ion Monitoring</i>	Detector Selectivo de Iones
SIMCA	<i>Soft Independent Modelling of Class Analogies</i>	Modelado Suave Independiente por Analogía de Clases
SMDI	-----	Sistema de Monitorización y Distribución de la Información
SNV	<i>Standard Normal Variate</i>	Variable Normal Estándar
SORS	<i>Spatially offset Raman</i>	Raman de Desplazamiento Espacial
SPC	<i>Statistical Process Control</i>	Control Estadístico del Proceso
SPME	<i>Solid-Phase Micro-Extraction</i>	Micro-extracción en Fase Dólide
SRS	<i>Spatially Resolved Spectroscopy</i>	Espectroscopia Espacialmente Resuelta
TCP/IP	<i>Transmission Control Protocol/ Internet Protocol</i>	Protocolo de control de transmisión/Protocolo de Red
TDD	<i>Total Daily Dose</i>	Dosis Máxima Diaria
TLC	<i>Thin Layer Chromatography</i>	Cromatografía en Capa Fina
TOC	<i>Total Organic Carbon</i>	Carbono Orgánico Total
UAB	-----	Universidad Autónoma de Barcelona
USP	<i>United States Pharmacopoeia</i>	Farmacopea de Estados Unidos
UV	<i>Ultraviolet</i>	Ultravioleta

RESUMEN

En esta Tesis Doctoral se han planteado, desarrollado e implementado soluciones basadas en la tecnología analítica de proceso (PAT) en distintos puntos del proceso de producción de medicamentos. Estas deben servir para encaminar la fabricación de medicamentos hacia una fabricación avanzada, dado que cada vez más las agencias reguladoras están incentivando a las compañías de sustancias activas, medicamentos de uso humano y de uso veterinario a fabricar en el paradigma PAT, QbD y fabricación en continuo con las continuas publicaciones de guías. Las técnicas analíticas rápidas, como son el NIR, el Raman o el IMS, junto a la quimiometría o tratamiento de datos multivariados son dos herramientas indispensables del PAT.

Los trabajos realizados se pueden clasificar en cuatro grandes bloques: métodos de identificación, análisis de proceso, validaciones de limpieza y análisis de datos históricos.

1. Para la identificación de materiales de partida que se formulan en los medicamentos se ha actualizado y validado una biblioteca espectral NIR que permite identificar y cualificar 224 sustancias empleando la estrategia de la biblioteca en cascada. La biblioteca principal se ha calculado mediante la distancia espectral, mientras que todas las sub-bibliotecas -en total 5 niveles- se han calculado con la distancia en el espacio PCA. También se ha realizado la evaluación inicial de un instrumento Raman, adquirido como técnica complementaria al NIR, para la identificación de materiales de partida. Se ha realizado un estudio que ha permitido detectar los factores o características inherentes a las muestras que podrían afectar a la señal Raman, lo que ha permitido establecer el protocolo del desarrollo de las bibliotecas correspondientes.

Dentro del bloque de identificación, se ha desarrollado e implementado un método NIR para la cualificación de comprimidos recubiertos en granel. Dado que la diferencia entre las presentaciones es física y no química -las tres dosificaciones tienen formulaciones proporcionales- ha sido necesario adquirir los espectros en transmisión y además, calcular el modelo con el análisis discriminante como algoritmo de reconocimiento de pautas sin aplicar ningún pretratamiento. De esta forma se ha obtenido un modelo específico y robusto.

2. Se ha implementado un método NIR que permite realizar el análisis en tiempo real de un producto farmacéutico granulado en el depósito de mezclado, una vez la operación unitaria de mezcla ha finalizado. Para ello se han calibrado y validado tres modelos -identificación,

cuantificación de API y contenido en agua- y se han llevado a cabo las etapas relacionadas con la industrialización de un sistema de medida PAT en una planta farmacéutica.

Se ha realizado el inicio de un estudio de viabilidad con un instrumento NIR que dispone de una sonda multipunto para implementarlo en la operación unitaria de dosificación de productos en polvo, granulados y comprimidos. Se ha demostrado que en las condiciones de trabajo, la señal NIR tiene una respuesta lineal frente a los cambios de concentración de API de dos productos, un granulado con un 5.0% de API y unos comprimidos con un 10.3%. No ha sido posible desarrollar un modelo PLS con los espectros de un medicamento que contiene un 1.0% de API ni con los de un producto en polvo que tiene la banda principal del API en una zona con una elevada presencia de ruido. Por otro lado, se ha demostrado la utilidad del instrumento para detectar de manera sencilla comprimidos rotos durante el proceso de dosificación.

3. La presencia de trazas de detergentes tras los procesos de limpieza de los equipos de producción se suele evaluar con técnicas analíticas no específicas. Se ha establecido una sistemática mediante diseño de experimentos que permite entender la influencia de los factores presentes en las muestras obtenidas durante la verificación de limpieza usando una técnica analítica no específica como lo es la medida de la conductividad del agua del último aclarado o la disolución resultante con la extracción del *swab*.

También se ha propuesto el uso de IMS como técnica específica para evaluar la presencia de detergentes. Se ha realizado un estudio de caracterización y cuantificación de surfactantes no iónicos –etoxilados- con IMS.

4. Durante la fabricación de un medicamento se generan un número elevado de datos por lo que se ha propuesto el uso de análisis multivariable con el que se puede demostrar el estado de control estadístico de los procesos, usando como ejemplo los datos relacionados con un comprimido recubierto. Para ello se han fijado los límites de control con los valores de Hotelling T^2 y los residuales Q calculados a partir de un análisis PCA de los atributos críticos de calidad y a partir de un *multiway* PCA para los parámetros de proceso que varían con el tiempo. Este último modelo puede usarse para analizar en tiempo real el proceso de producción.

OBJETIVOS

OBJETIVOS

El objetivo principal de esta Tesis Doctoral es aportar soluciones basadas en la tecnología analítica de procesos (PAT), que permitan evolucionar la fabricación tradicional de productos farmacéuticos hacia una fabricación avanzada. Las técnicas analíticas rápidas, como son el NIR, Raman o IMS, y la quimiometría o tratamiento de datos multivariados, son dos herramientas indispensables de PAT.

De manera más detallada, la Tesis pretende desarrollar herramientas PAT en las diferentes operaciones unitarias que están implicadas en la fabricación de un medicamento, como son:

- La identificación y cualificación de los materiales de partida, a su recepción.
- La determinación de distintos atributos críticos de calidad (CQA) durante la producción de los medicamentos que permita tener un control en tiempo real.
- Proponer mejoras en el proceso de validación y verificación de los procesos de limpieza, para obtener la máxima información con los mínimos tiempo y coste posibles.
- Demostrar la importancia de los datos históricos de parámetros de calidad de los materiales de partida, de parámetros de proceso y de atributos de calidad de los productos intermedios y terminados, para el proceso de mejora continua y para el desarrollo de los procesos existentes bajo las premisas de la calidad por diseño.

Parte I

INTRODUCCIÓN

1. INDUSTRIA FARMACÉUTICA AVANZADA

1.1. ENTORNO

La tesis se ha realizado en la industria farmacéutica, en la compañía Laboratorios Menarini, S.A. ubicada en Badalona (España). La empresa está presente en España desde el año 1961, cuando el pequeño laboratorio español Puig Sala se fusiona con la empresa italiana A. Menarini IFR. Actualmente, la planta tiene una superficie de 15000m² y está formada por 750 trabajadores. De media se fabrican 70 millones de unidades al año, de las cuales aproximadamente el 80% son medicamentos sólidos orales.

Desde hace más de 20 años, Laboratorios Menarini ha apostado por la innovación en el campo de la fabricación de medicamentos, como demuestra su convenio de colaboración con el grupo de investigación de Quimiometría Aplicada de la Universidad Autónoma de Barcelona. Durante todo este tiempo se han desarrollado un gran número de métodos para el control de calidad mediante el uso de técnicas analíticas rápidas. También se han llevado a cabo estudios de viabilidad de calidad por diseño (QbD del inglés *Quality by Design*) y de la aplicación de tecnología analítica de proceso (PAT del inglés *Process Analytical Technology*). Estas herramientas son imprescindibles para un nuevo paradigma de fabricación, la fabricación en continuo (CM del inglés *Continuous Manufacturing*). Ahora bien, la falta de expertos dentro de la empresa ha complicado la aplicación de dichas innovaciones en rutina. Se concluyó que era imprescindible empezar a desarrollar los proyectos desde dentro de la compañía, aprovechando el conocimiento aportado por la universidad. Surgió la posibilidad de realizar la presente tesis mediante una beca de Doctorado Industrial, otorgada por el organismo AGAUR de la Generalitat de Catalunya (expediente 2015-DI 014). Precisamente, el objetivo de estos doctorados es el de transferir el conocimiento desde las universidades a las industrias.

La fabricación de un medicamento consta de múltiples etapas, como son la recepción de la materia prima, pasando por todas las operaciones unitarias implicadas en la fabricación del medicamento o el control de calidad del mismo. Hay muchos departamentos implicados para que un producto terminado esté en el mercado, entre los que se encuentran garantía de calidad, tecnología farmacéutica, producción, control de calidad, ingeniería o mantenimiento.

La industria farmacéutica es un sector altamente regulado por las agencias competentes en cada uno de los países. Un medicamento solo se puede comercializar en un país si la agencia pertinente ha aprobado el registro del mismo. Desde hace tiempo las agencias reguladoras europea y americana, EMA (*European Medicine Agency*) y FDA (*Food and Drug Administration*), incentivan

a la industria a hacer un cambio en la forma de fabricación, esto es, adoptar el desarrollo de procesos de fabricación mediante QbD y a controlar la calidad del proceso y de los productos con el uso del PAT. Además, se está planteando cambiar la fabricación tradicional por lotes por la fabricación en continuo. Implícitamente, para poder aplicar PAT, también es necesaria la implementación de técnicas analíticas rápidas que permitan la obtención de resultados en tiempo real.

1.2. FABRICACIÓN TRADICIONAL DE UN MEDICAMENTO SÓLIDO

La fabricación de productos farmacéuticos abarca desde la adquisición de materiales, la producción, el control de calidad, la liberación, el almacenamiento hasta la distribución del propio medicamento y los controles correspondientes a todas las operaciones mencionadas¹. Esta fabricación se debe realizar bajo las normas de correcta fabricación (NCF o GMP del inglés *Good Manufacturing Practices*).

La fabricación de un medicamento consta de distintas operaciones unitarias, donde en cada una de ellas hay un cambio químico o físico del material². En la Figura 1.1 se muestran ejemplos de diagramas de flujo de las operaciones unitarias implicadas para la fabricación de algunas de las fórmulas sólidas orales más fabricadas, como son comprimidos, polvos o granulados en sobres y cápsulas. La forma tradicional de proceder es que dichas operaciones sean discretas, esto es, cierta cantidad de material se carga en el equipo y una vez finalizado el proceso, se descarga el producto (ya sea producto intermedio o producto granel). En este caso, se entiende como lote una cantidad

de producto fabricado que parte de los mismos materiales de partida y ha pasado por las mismas operaciones unitarias, considerándose homogéneo¹. La calidad del producto se garantiza por medio de ensayos analíticos en los materiales de partida, un proceso de fabricación invariable, controles en proceso y análisis de los productos intermedios, graneles y/o terminados. Una vez el producto terminado cumple las especificaciones, este se liberará al mercado.

Las etapas que se siguen para que un nuevo producto salga al mercado son las siguientes:

- Definición del producto: definir el producto objetivo así como los atributos de calidad del mismo, lo que se conoce como perfil de la calidad objetivo del medicamento (QTPP del inglés *Quality Target Product Profile*).

Desarrollo procesos: una vez definido el producto objetivo, se empieza a desarrollar el proceso de producción. El objetivo de esta etapa es que queden definidos los atributos de calidad del material de partida, además de fijar los puntos de consigna y rangos de operación de los parámetros de proceso. Esto se suele llevar a cabo mediante el *know-how* de la empresa y realizando experimentos univariantes³. Además, mediante el conocimiento previo y realizando un análisis de riesgos se determina, de entre todos los parámetros de calidad, cuales son críticos (CQA del inglés *Critical Quality Attributes*).

También forma parte de la etapa de desarrollo establecer una estrategia de control. En el enfoque tradicional esta se suele establecer para demostrar que el producto intermedio (producto obtenido en una operación unitaria intermedia), granel o terminado cumple con las especificaciones establecidas. Para ello se suelen usar técnicas analíticas que se encuentran instaladas de manera *off-line*; esto es, en unos laboratorios apartados de la planta de producción, por lo que se realiza una vez ha finalizado el proceso que está siendo evaluado.

- Validación del proceso: todo proceso utilizado para producir medicamentos que se comercializan en el mercado debe haber sido validado siguiendo los principios del Anexo 15 de las GMP. El objetivo de validar un proceso es establecer pruebas documentales de que el proceso de fabricación desarrollado es capaz de producir de manera consistente el producto con una calidad determinada⁴. El enfoque tradicional de validación se basa en demostrar que, con los parámetros de calidad del material de partida, las cantidades de material establecidos y los rangos de los parámetros de operación (definidos durante el desarrollo), se obtiene el producto dentro de las especificaciones predefinidas en el protocolo (atributos de calidad) de una manera constante. Para ello se fabrica un número determinado de lotes de producto en las condiciones de rutina para confirmar la reproducibilidad del mismo⁵.

INTRODUCCIÓN

- Registro: solo se puede comercializar un producto en el mercado si ha sido aprobado por la agencia regulatoria competente del país. Para este fin, se debe desarrollar un expediente si el producto es nuevo o se debe modificar uno existente si se pretenden cambiar características de un producto autorizado. Una vez aprobado, el producto se puede fabricar en rutina.
- Fabricación en rutina: una vez autorizados, los productos se fabrican siguiendo lo definido en el expediente y bajo un sistema de calidad. Además, se debe revisar periódicamente la calidad de cada producto para detectar cualquier tendencia, e identificar mejoras de proceso y/o producto⁶, elaborando informes periódicos de calidad (PQR del inglés *Product Quality Review*).

El enfoque tradicional de fabricación de medicamentos presenta ciertas desventajas. Entre otras, se encuentran:

- No se dispone de un conocimiento profundo del sistema. Esto conlleva que cuando hay resultados fuera de especificaciones (OOS del inglés *Out Of Specification*) sea muy complicado encontrar las causas raíces que lo han provocado. Esto es, no hay una base científica para entender las relaciones entre los parámetros de partida –calidad del material de partida y parámetros de proceso- con la calidad del producto procesado.
- Durante el desarrollo, se establecen parámetros de calidad de materia prima y de proceso invariables o con rangos muy estrechos. Esto provoca que se deseche producto debido al poco margen operacional.
- La cantidad de material que se fabrica por lote es invariable, con lo que no se puede disminuir o aumentar según la demanda que hay en cada momento.
- Al establecerse una estrategia de control con un elevado consumo de tiempo, hay una acumulación de existencias. Esto es, el producto intermedio no puede pasar a la siguiente operación unitaria o el producto terminado no se puede aprobar, si antes no ha sido evaluado de conformidad por el departamento de control de calidad. Otro problema es que si se detecta algún defecto una vez finalizada la fabricación del medicamento, el lote se debe reprocesar, recuperar o desechar⁷.
- El PQR suele realizarse con análisis univariantes de los datos; lo que esconde fuera de control o tendencias, que saldrían a la luz realizando análisis multivariantes. Esto contribuiría a mejorar la comprensión del proceso de fabricación, ya que podría ayudar a identificar la causa raíz de OOS, y a entender relaciones entre variables.

1.3. ENFOQUE MEJORADO DE LA FABRICACIÓN DE MEDICAMENTOS

Distintas causas justifican la necesidad de implementar un nuevo enfoque de fabricación en la industria farmacéutica, tanto desde el punto de vista de la industria, como desde el punto de vista de las agencias reguladoras.

Las compañías farmacéuticas están interesadas en reducir los costes de producción debido al incremento de coste de la investigación y desarrollo y la presión de los gobiernos por reducir sus precios de los medicamentos⁸. También es conveniente aumentar la flexibilidad de la definición y tamaño de los procesos, ya que el futuro de los medicamentos pasará por desarrollar terapias individualizadas que requerirán de lotes de tamaño variables⁹. Esto provocará que sea necesario disponer de sistemas que permitan tener un amplio portafolio de productos y además, que un mismo proceso sirva para fabricar distintos volúmenes de producto y así abastecer según la demanda del mercado⁹. Otro factor a tener en consideración es que se está viviendo lo que algunos autores definen como la cuarta revolución industrial o Industria 4.0¹⁰⁻¹². Se está transformando la industria 3.0, la cual se caracterizó por computarizar y automatizar los sistemas, a la industria 4.0, que se basa en la implementación de sistemas inteligentes y autónomos mediante el uso de los datos que se generan y el aprendizaje automático (*machine learning*). Los elementos centrales de esta nueva industria son el internet de las cosas (IoT del inglés *Internet of Things*), internet de los servicios (IoS, del inglés *Internet of Services*), sistemas cibernéticos (CPS, del inglés *Cyber-Physical Systems*) y Smart Factory¹³. Estos ayudarán a optimizar la cadena de valores, a reducir costes y a ahorrar energía¹³.

Por otro lado, las agencias reguladoras están interesadas en que las compañías dispongan de fabricaciones eficientes, ágiles y flexibles, que permitan obtener productos de elevada calidad sin una supervisión regulatoria extensa¹⁴ y así solventar el problema de escasez de suministros (*drug shortage*). Las EMA y FDA, junto al Consejo Internacional para la armonización de los requisitos técnicos para el registro de medicamentos de uso humano (ICH del inglés *International Council for Harmonisation*), han fomentado distintas iniciativas para fabricar en un nuevo paradigma. Prueba de ello son las normativas y guías publicadas en este sentido^{4,5,15-21}.

El nuevo enfoque de fabricación es lo que se conoce como fabricación en continuo, desarrollando los procesos mediante QbD y maximizar tanto como sea posible el control de calidad con PAT.

1.3.1. Fabricación en continuo

La fabricación en continuo consiste en un proceso de fabricación sin interrupciones, esto es, el sistema se alimenta con material de partida que va fluyendo por las distintas operaciones unitarias y se descarga producto acabado de manera continua. Además, se asegura la calidad del mismo en

tiempo real^{9,22}. En este caso, se define como lote la cantidad de material producido en un periodo de tiempo determinado¹ y que cuenta con características uniformes y con la calidad conforme a especificaciones²².

La fabricación en continuo se caracteriza por tener un proceso integrado con un número reducido de operaciones unitarias separadas y por usar equipos de tamaño reducido. A la vez, debe contar con un sistema de ingeniería bien diseñado, para adaptar de manera automática el proceso según la variación de las propiedades de los materiales y a las condiciones ambientales. El desarrollo del proceso se realiza mediante un enfoque mejorado, esto es, el uso de QbD y PAT permiten obtener información de la calidad del producto en tiempo real⁹. Algunas de las ventajas que presenta este tipo de fabricación en la industria farmacéutica son:

- Aumento de la robustez del proceso, debido a que se tiene un conocimiento profundo del mismo y se controla de manera proactiva. Los equipos usados en la fabricación son los mismos que durante la fase de desarrollo del proceso, por lo que todo el conocimiento adquirido se puede aplicar durante la fabricación²³. Esto influye en un aumento de la calidad del producto y en una reducción de los costes por la generación de rechazos.
- Flexibilidad de producción. El volumen de producto fabricado puede aumentar o disminuir variando el tiempo de producción, para adaptarse a las demandas⁹.
- Al usarse maquinaria de tamaño reducido, se puede transportar a cualquier lugar, por lo que se puede llegar a una descentralización de la fabricación. También se reduce el impacto ambiental⁹

Como se ha comentado anteriormente, la fabricación en continuo está altamente ligada al QbD y al PAT. Por lo que a continuación se describirán estos dos conceptos.

1.3.2. Calidad por diseño (QbD)

La ICH publicaba en 2004 la primera guía ICH Q8, donde proponía el QbD como herramienta para desarrollar productos farmacéuticos. En ella se define el QbD como un enfoque sistemático para el desarrollo, que comienza con unos objetivos predefinidos y da importancia a la comprensión y control del proceso y del producto, basándose en una ciencia sólida y en la gestión de riesgos de la calidad¹⁸. Este enfoque parte de la base que la calidad se debe construir durante el diseño¹⁸. A partir de ese momento, se han publicado nuevas guías dando soporte a esta idea, como son la ICH Q9, Q10 y Q11. Actualmente la ICH Q12 sobre consideraciones técnicas y regulatorias de la gestión del ciclo de vida de los medicamentos, está publicada en *draft*. Además hay dos *concept papers* de la ICH Q13 y Q14 en relación a la fabricación en continuo de API y medicamentos y al desarrollo de métodos analíticos adaptados a la nueva realidad, respectivamente.

Los objetivos de QbD son, entre otros, incrementar la capacidad del proceso, la robustez y la efectividad de fabricación. Mejorar la capacidad y velocidad de desarrollo de nuevos productos y/o procesos. Aumentar la flexibilidad regulatoria y agilizar la gestión de cambios post aprobación dentro del marco de mejora continua. Por último, mejorar el análisis de causas raíces y disminuir tanto como sea posible el análisis del producto terminado^{2,7,24}.

Los pasos a seguir durante el desarrollo de un producto por QbD se enumeran a continuación:

- 1) Definir el QTPP donde se detallan todas las características de calidad que debe tener el medicamento.
- 2) Identificar qué características del producto son críticas (CQAs del inglés *Critical Quality Attributes*), mediante el uso de conocimiento previo y del análisis de riesgos².
- 3) Diseñar el producto y el proceso para obtener un elevado conocimiento de los mismos. En esta etapa se deben identificar los atributos críticos de material (CMA del inglés *Critical Material Attributes*), las operaciones unitarias críticas y los parámetros críticos del proceso (CPP del inglés *Critical Process Parameters*)². Para este fin, se usa el conocimiento previo, la experimentación, el análisis de riesgos y el conocimiento científico que permite identificar y entender el impacto en los atributos críticos de calidad^{3,18}. Esto servirá para encontrar las relaciones entre CMAs y CPPs con los CQAs del producto². El proceso se entiende profundamente cuando se han detectado todas las fuentes de variabilidad, el proceso gestiona esta variabilidad y los atributos de calidad del producto se pueden predecir de manera exacta y fiable².
- 4) Si es necesario, se establece el espacio de diseño del proceso²⁴. Este espacio de diseño debe estar localizado dentro de los límites de todo el espacio de conocimiento adquirido durante el desarrollo del proceso.
- 5) Definir una estrategia de control, tanto de los materiales de partida, del proceso, y si es necesario, del producto terminado². El objetivo es asegurar que el producto terminado cumplirá con los estándares de calidad establecidos^{18,24}. Se definen tres niveles de estrategia de control²:
 - El primero es aquel que permite la monitorización de las CQAs en tiempo real. Las CMAs se monitorizan y los parámetros de proceso se ajustan automáticamente para asegurar la calidad del producto. Este nivel permitiría la liberación en tiempo real, esto es, la calidad del producto quedaría asegurada durante el proceso.
 - El segundo nivel consiste en controlar el producto con una reducción de análisis del producto terminado y con la flexibilidad de tener los CMAs y CPPs dentro del espacio de diseño.

- El tercer nivel consiste en el control tradicional de productos farmacéuticos. Se realiza un amplio número de análisis para cada producto terminado. Asimismo, hay un bajo margen de aceptación de rangos de CMA y CPPs.

Se entiende que para la aplicación QbD se usa el primer nivel, el segundo o bien una combinación de ambos niveles². También será necesario en estos dos casos establecer una monitorización regular del funcionamiento de los modelos predictivos para asegurar su validez¹⁸.

- 6) Una vez el proceso está en rutina, se propone la mejora continua del proceso y/o del producto y la verificación continua del proceso. Se monitoriza el producto y el proceso, se analizan y evalúan los datos para asegurar que el proceso sigue funcionando dentro del espacio declarado. También se debe demostrar que el proceso sigue en estado de control estadístico y que el proceso es capaz^{2,7}. Si se trabaja con un espacio de diseño, es importante monitorizar regularmente el buen funcionamiento del modelo¹⁸.

Las herramientas disponibles para llevar a cabo las etapas comentadas anteriormente son el análisis de riesgo, analizadores de proceso, diseño de experimentos, análisis multivariable y modelado de proceso^{7,24}.

Este nuevo enfoque se puede introducir en cualquier punto del ciclo de vida de un medicamento. No solo en la etapa de desarrollo de un nuevo medicamento, sino también en la etapa de su fabricación rutinaria.

1.3.3. Tecnología Analítica de Proceso (PAT)

PAT se define como un sistema para diseñar, analizar y controlar el proceso de fabricación mediante medidas en tiempo real de los atributos críticos de calidad de materiales de partida, atributos de calidad de productos intermedios y parámetros de proceso, para asegurar la calidad del producto terminado²¹. Se debe entender el PAT como un sistema que requiere tomar acciones basadas en la información que se obtiene en tiempo real⁷.

El PAT abarca un conjunto de herramientas y principios que permiten mejorar el conocimiento del proceso de fabricación y su control²⁵. Se consideran herramientas PAT la quimiometría o análisis multivariable de datos, analizadores de proceso, herramientas de control y automatización de proceso, herramientas de mejora continua y gestión del conocimiento²¹. En la Tabla 1.1 hay una descripción y ejemplos de cada una de ellas.

Tabla 1.1 Herramientas para aplicar sistemas PAT²¹.

Herramientas	Definición	Ejemplos
Quimiometría	Herramientas matemáticas y estadísticas para el diseño, la adquisición y el análisis de datos	<ul style="list-style-type: none"> - Diseño de experimentos y superficie de respuesta - Simulación de proceso - Reconocimiento de pautas - Modelos de predicción
Sensores y analizadores de proceso	Herramientas para medir atributos químicos, físicos o microbiológicos a tiempo real del proceso	<ul style="list-style-type: none"> - Sensores para medidas univariantes (pH, Temperatura, presión...) - Analizadores para medidas multivariantes (NIR, UV, Raman,...)
Herramientas de control de proceso	Herramientas que permitan monitorizar el estado del proceso en tiempo real analizando los datos obtenidos y, si es necesario, tomar decisiones de manera automática	<ul style="list-style-type: none"> - Sistemas de monitorización y distribución de la información
Herramientas de mejora continua y gestión del conocimiento	Herramientas para analizar todos los datos obtenidos durante el ciclo de vida de un producto para aumentar el conocimiento del mismo y del proceso; y si se considera necesario, proponer mejoras	<ul style="list-style-type: none"> - PQR multivariable

Los sensores y analizadores del proceso se pueden encontrar instalados de las siguientes maneras^{7,21}:

- *At-line*: la muestra debe extraerse de la operación unitaria y medir con un analizador que se encuentra instalado cerca del proceso.
- *On-line*: la operación unitaria tiene incorporado un sistema que deriva la muestra hacia el analizador y esta puede ser devuelta al proceso o no.
- *In-line*: el analizador está integrado en la operación unitaria de manera invasiva o no.

La aplicación del PAT en la fabricación comercial se puede usar para dos objetivos:

- 1) Como herramienta de recopilación de conocimiento para respaldar la robustez del proceso (concepto de “firma de proceso”). Esto permite la mejora continua y así resolver problemas de calidad y robustez³.
- 2) Como estrategia de control sustituyendo los controles tradicionales *off-line*.

Debido a que la información se obtiene en tiempo real, la implementación de un sistema PAT podría llegar a permitir liberar producto en tiempo real (RTRT del inglés *Real Time Release Testing*); si se combina con un profundo conocimiento del proceso (el concepto QbD). Se acaba de

publicar un anexo de las GMP específico para el RTRT. En ella se define RTRT como “una combinación de controles y monitorizaciones en proceso que puede reemplazar, cuando esté autorizado, a los ensayos en el producto final como parte de la decisión de liberación de un lote”²⁶.

Aunque hace años ya desde la primera publicación de la guía PAT y durante todo este tiempo se hayan publicado todas las guías mencionadas respecto al QbD o al cambio de paradigma de fabricación, mayoritariamente aún se fabrica de la manera tradicional. Varios motivos pueden influir a que el cambio de paradigma de fabricación sea un proceso lento. Uno de ellos puede ser que la inversión inicial que se debe efectuar tanto para estudios de viabilidad como para la instalación de equipos y sistemas que permitan el cambio es elevada. Otro puede ser la falta de recursos humanos que tengan el conocimiento en este campo. Finalmente, debido que vivimos en un mundo globalizado, las farmacéuticas suelen vender en un número elevado de mercados. Aunque la FDA y la EMA estén a favor de estos cambios, la falta de confianza a que otras agencias reguladoras acepten dichos cambios, podría frenar la apuesta de las compañías a invertir en este sentido. También cabe destacar las diferencias que existen entre las compañías farmacéuticas instaladas en EEUU, donde el número de aplicaciones PAT, QbD y fabricación en continuo parece que es mayor que las que se llevan a cabo en Europa.

1.4. BIBLIOGRAFÍA

1. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Glosario. (2010).
2. Yu, L. X. *et al.* Understanding Pharmaceutical Quality by Design. *AAPS J.* **16**, 771–783 (2014).
3. Simon, L. L. *et al.* Assessment of recent process analytical technology (PAT) trends: A multi-author review. *Org. Process Res. Dev.* **19**, 3–62 (2015).
4. ICH (International Conference for Harmonisation). ICH Q11 on development and manufacture of drug substances. (2012).
5. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Anexo 15 Cualificación y validación. (2015).
6. AEMPS. Guía de NCF de Medicamentos de Uso Humano y Veterinario, capítulo 1: Sistema de calidad farmacéutico. 1–7 (2013).
7. Schlindwein, W. S. & Gibson, M. *Pharmaceutical Quality by Design. A practical Approach.* (Wiley, 2018).
8. Bostijn, N., Van Renterghem, J., Dhondt, W., Vervaeke, C. & De Beer, T. A continuous manufacturing concept for a pharmaceutical oral suspension. *Eur. J. Pharm. Sci.* **123**, 576–583 (2018).
9. Kleinebudde, P., Khinast, J. & Rantane, J. *Continuous Manufacturing of Pharmaceuticals.* (John Wiley & Sons Ltd., 2017).
10. Duarte, A. Y. S., Sanches, R. A. & Dedini, F. G. Assessment and technological forecasting in the textile industry: From first industrial revolution to the Industry 4.0. *Strateg. Des. Res. J.* **11**, 193–202 (2018).

11. Shvetsova, O. A. & Kuzmina, A. D. Development of Engineering Personnel in the Era of the Fourth Industrial Revolution. in *3rd International Conference Ergo-2018: Human Factors in Complex Technical Systems and Environments* 45–48 (2018).
12. Griffiths, F. & Ooi, M. The fourth industrial revolution-Industry 4.0 and IoT [Trends in Future I&M]. *IEEE Instrum. Meas. Mag.* **21**, 29–43 (2018).
13. Ding, B. Pharma Industry 4.0: Literature review and research opportunities in sustainable pharmaceutical supply chains. *Process Saf. Environ. Prot.* **119**, 115–130 (2018).
14. Yu, L. X. & Kopcha, M. The future of pharmaceutical quality and the path to get there. *Int. J. Pharm.* **528**, 354–359 (2017).
15. EMA. Guideline on Real Time Release Testing (formerly Guideline on Parametric Release). (2009).
16. ICH (International Conference for Harmonisation). ICH Q10 - Pharmaceutical quality system. (2008).
17. ICH (International Conference for Harmonisation). ICH Q9 - Quality Risk Management. (2005).
18. ICH (International Conference for Harmonisation). ICH Q8 (R2) - Pharmaceutical Development. (2009).
19. ICH (International Conference for Harmonisation). ICH Q12 - Technical and Regulatory Considerations for Pharmaceutical Product Lifecycle Management (Draft). (2017).
20. EMA. Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations. *Eur. Med. Agency* **44**, 1–28 (2014).
21. FDA. Guidance for Industry PAT: A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. 16 (2004).
22. Lee, S. L. *et al.* Modernizing Pharmaceutical Manufacturing: from Batch to Continuous Production. *J. Pharm. Innov.* **10**, 191–199 (2015).
23. Moghtadernejad, S. *et al.* A Training on: Continuous Manufacturing (Direct Compaction) of Solid Dose Pharmaceutical Products. *J. Pharm. Innov.* **13**, 155–187 (2018).
24. ISPE (International Society for Pharmaceutical Engineering). Part 1-Product Realization using Quality by Design (QbD): Concepts and Principles. in *Product Quality Lifecycle Implementation (PQLI) from concept to Continual Improvement* 184 (2011).
25. Dickens, J. E. Chapter 1 - Overview of Process Analysis and PAT. in *Process Analytical Technology* (ed. Katherine, A.) 1–15 (2010).
26. AEMPS. Guía de NCF de Medicamentos de Uso Humano y Veterinario. Anexo 17: Ensayos de liberación a tiempo real y liberación paramétrica. 9 (2018).

2. HERRAMIENTAS DE OBTENCIÓN DE DATOS

Los datos usados para el desarrollo de esta tesis provienen principalmente de dos fuentes. La primera, de técnicas analíticas rápidas como son la espectroscopia del infrarrojo cercano (NIR del inglés *Near-Infrared*), la espectroscopia Raman y la espectrometría de movilidad iónica (IMS del inglés *Ion Mobility Spectrometry*). La segunda fuente son todos los datos que se obtienen relacionados con la fabricación de un producto; incluyendo los atributos de los materiales de partida, los datos de los procesos y los atributos de calidad de los productos intermedios y graneles. A continuación se describe cada una de las herramientas empleadas.

2.1. ESPECTROSCOPIA NIR

La espectroscopia NIR es una técnica de absorción de la radiación cuando esta interacciona con la materia y produce cambios en los estados vibracionales de las moléculas. El descubrimiento de la región espectral NIR se atribuye al astrónomo Sir William Herschel, pero no es hasta mediados del siglo XX que se empiezan a desarrollar aplicaciones industriales^{1,2}.

2.1.1. Fundamentos de la radiación NIR

La energía asociada a las bandas de una molécula determinada es la suma de las energías electrónicas, vibracionales y rotacionales. Para cada nivel electrónico, hay varios niveles de vibración y a su vez varios estados rotacionales. La energía de la radiación de la región IR produce transiciones vibracionales y rotacionales³. Para entender las transiciones utilizamos una molécula diatómica. Su vibración se puede describir como el de un oscilador anarmónico, tal como se muestra en el diagrama de energía potencial de la Figura 2.1, donde los niveles de energía están cuantizados. Al tratarse de un oscilador anarmónico, los niveles energéticos no son equidistantes, sino que disminuyen al aumentar v . Esta propiedad provoca que las reglas de selección no solo

INTRODUCCIÓN

permiten transiciones entre niveles contiguos ($\Delta v = \pm 1$), sino que también están permitidas entre estados vibracionales con $\Delta v = \pm 2, \pm 3 \dots$ ^{1,4}. El modelo para una molécula con un número mayor de átomos es más complejo, pero la esencia es la explicada anteriormente.

La radiación IR ocupa la región que comprendida entre 10 y 12820 cm^{-1} del espectro electromagnético, que a su vez se puede dividir en 3 subregiones que provocan transiciones distintas (Tabla 2.1).

Tabla 2.1 Descripción de las tres regiones del IR.

Regiones del IR	Rango (cm^{-1})	Transiciones que se observan
FIR (infrarrojo lejano)	10 a 400 cm^{-1}	- Rotaciones puras - Vibraciones en red de sólidos ³
MIR (infrarrojo medio)	400-4000 cm^{-1}	- Bandas fundamentales (transiciones del estado vibracional fundamental al primero excitado ($v=0 \rightarrow v=1$))
NIR (infrarrojo cercano)	4000-12800 cm^{-1}	- Sobretonos (transiciones vibracionales entre estados $\Delta v > 1$) - Combinación de bandas fundamentales (un fotón excita simultáneamente dos o más modos de vibración)

Dado que las transiciones entre estados con $\Delta v > 1$ son menos probables, las bandas que se obtienen en el NIR son menos intensas que en el MIR. También cabe mencionar que dentro de los sobretonos el primer ($v=0 \rightarrow v=2$) es más intenso que el segundo ($v=0 \rightarrow v=3$) y así sucesivamente. Hay otros efectos presentes en los espectros NIR como son la resonancia de Fermi (característica del CO_2), la de Darling-Dennison (característica de la molécula de H_2O) y modelos locales^{1,2}.

Para que una molécula sea activa en el NIR, se deben cumplir tres condiciones. La primera es que, como en cualquier técnica espectroscópica de absorción, la energía del fotón incidente debe coincidir con la diferencia energética de los niveles, en este caso, vibracionales. La segunda es que, como en todas las técnicas IR, la vibración de la molécula debe producir un cambio en el momento dipolar de la misma. Por último, los átomos implicados en la vibración deben presentar una anarmonicidad de enlace -diferencia en las masas atómicas- elevada. Así, los grupos más activos serán los X-H (X=C, N, O o S)². Aparte de los enlaces moleculares, también se obtiene información de interacciones intermoleculares como son los puentes de hidrógeno.

Los espectros NIR se caracterizan por presentar bandas anchas y superpuestas. Esto hace que la selectividad de la técnica sea baja, razón por la cual se emplea en combinación con la quimiometría. Por otro lado, las intensidades de absorción son débiles, debido a la baja probabilidad de que ocurra el fenómeno que da lugar al espectro NIR, lo que hace que la sensibilidad de la técnica sea baja. La mejora instrumental ha mejorado dicha sensibilidad⁵. El

hecho que tenga baja absorptividad, se convierte en una ventaja debido a que permite medidas directas sin preparación de muestra y mayor penetración de radiación en las muestras sólidas⁶. Esta combinación junto con otras ventajas que presenta la técnica (Tabla 2.2) ha hecho que la implementación del NIR en la industria farmacéutica vaya en aumento.

Tabla 2.2 Ventajas de la técnica espectroscópica NIR.

Es una técnica no invasiva ni destructiva, por lo que permite análisis <i>in-</i> y <i>on-line</i>
Registro de espectros rápido, lo que permite el análisis en tiempo real
De un mismo espectro se puede obtener información física y química de la muestra
Versátil, permite registrar el espectro de muestras sólidas, líquidas y gaseosas
Ningún pretratamiento de la muestra o mínimo, por lo que además de contribuir a poder realizar análisis <i>in-</i> y <i>on-line</i> , disminuye el error asociado a la preparación de la muestra, y es respetuoso con el medioambiente
Posibilidad de uso de fibras ópticas (>100m)
Coste de análisis relativamente bajo

2.1.2. Instrumentación

Un instrumento NIR consta básicamente de cuatro partes: una fuente de radiación, un sistema de selección de longitudes de onda, un compartimento o la interfaz de contacto con la muestra y un detector. También es necesario un convertidor de la señal.

2.1.2.1. Fuente de radiación

La fuente de radiación más usada es la lámpara de halógeno con filamento de tungsteno y ventana de cuarzo. Esta emite una radiación policromática, de energía alta y constante en todo el rango NIR y también se caracterizan por ser estables⁷. Las trazas de halógeno sirven para reciclar el tungsteno vaporizado volviéndolo a depositar al filamento y así alargar la vida de la lámpara^{8,9}. Otro tipo de lámparas que se pueden usar son las LED. Excepto estas últimas, el resto de fuentes de radiación necesitan un selector de longitudes de onda⁸.

2.1.2.2. Sistema de selección de longitudes de onda

El objetivo es filtrar el haz de luz policromática en valores discretos de longitudes de onda⁹ de una anchura de banda estrecha, exacta y precisa respecto a la banda de absorción que mide y de una elevada intensidad en todo el rango de longitudes de onda¹⁰. Este sistema puede ir antes o después de la muestra. Hay distintos sistemas de selección: monocromadores, policromadores, filtros, AOTF e interferómetros. Se hace una pequeña descripción y se mencionan algunas de las características más importantes en la Tabla 2.3. La elección del sistema se hará según la aplicación que se vaya a desarrollar.

Tabla 2.3 Sistemas de longitudes de onda disponibles para equipos NIR y algunas características.

Equipo	Sistema de selección	Descripción	Propiedades
Dispersivo	Monocromadores de red	Realizan barridos espectrales difractando la luz en diferentes ángulos ⁹ . Consta de dos rendijas, una de entrada y otra de salida, y de una red de difracción móvil. Las longitudes de onda se registran sucesivamente en el tiempo.	+ Obtención de todo el rango espectral + Alta relación S/N - Baja sensibilidad - Alto tiempo de adquisición
	Policromadores	Consta de una rendija de entrada, una red de difracción fija y una red de diodos. La red de diodos realiza el barrido espectral permitiendo la detección simultánea de longitudes de onda con varios canales ⁶ .	+ Obtención de todo el rango espectral + Velocidad de adquisición elevada + Miniaturización (no tiene partes móviles) y robustos - Baja precisión - Posible calentamiento de la muestra
No dispersivo	Filtros	Consta de un filtro o un seguido de ellos, los cuales actúan como una ventana que o bien deja pasar una parte del espectro o bien bloque todas las longitudes de onda por encima o por debajo de cierta frecuencia ¹¹ .	+ Compactos y resistentes - Solo se obtienen ciertas longitudes de onda
	Filtros acústico-óptico (AOTF)	Constan de materiales birrefringentes (TeO ₂) con uno o varios transductores piezoeléctricos. Las longitudes de onda se escogen usando señales de radio-frecuencia, las cuales alteran el índice de refracción del TeO ₂ ⁴ .	+ Se pueden miniaturizar (no tiene partes móviles) + Robusta + Repetitividad - Baja estabilidad y reproducibilidad
	Transformada de Fourier	La radiación es modulada mediante el interferómetro permitiendo la codificación de la longitudes de onda y la medida de longitudes de onda se realiza simultánea ⁶ .	+ Alta precisión y reproducibilidad de longitud de onda + Elevada velocidad de adquisición - En algunos casos pueden ser sensibles a la vibración

2.1.2.3. Compartimento de la muestra

El compartimento de la muestra o interfase entre el equipo y la muestra vendrá determinado por dos factores, el modo de adquisición del espectro y la aplicación de la técnica.

Existen tres modos de adquisición del espectro, reflectancia, transmisión y transflectancia. En la Figura 2.2 se presenta un esquema de cada uno de ellos.

- Reflectancia difusa: la luz incidente es absorbida por la muestra o reflejada. Esta reflexión puede ser especular -mismo ángulo que la radiación incidente sin haber interactuado con la muestra, por lo que no contiene información de la misma- o difusa -la radiación que interactúa con la muestra y que no es absorbida por la misma, es dispersada en todas direcciones. Se mide la radiación dispersada (*scattering*) que vuelve al detector, el cual está situado al mismo lado que la fuente de radiación.
- Transmisión: se mide la radiación que traspasa la muestra sin haber sido absorbida y llega al detector localizado en lado opuesto de la fuente de radiación.
- Transflectancia: se coloca un reflector después de la muestra. La radiación que no ha sido absorbida por la muestra es devuelta al detector que se encuentra en el mismo lado que la fuente de radiación.

Otro factor a tener en cuenta para elegir el compartimento de muestra es la forma de aplicación de la técnica, esto es, si se pretende usar en el laboratorio (*off-line*) o en el proceso de fabricación (*off-, at-, in- u on-line*).

Teniendo en cuenta esto, en el mercado hay disponible una gran variedad de compartimentos o interfaces con las muestras entre ellos se encuentran:

- Esferas integradoras para medir muestras en modo reflectancia difusa. Estas se usan en *at-* u *off-line*.
- Porta cubetas para medir líquidos en modo transmitancia en *at-* u *off-line*.
- Sondas con fibra óptica (tanto de transmisión como de reflectancia difusa) para medidas *off-, at-, in-* u *on-line* de muestras sólidas y líquidas.

2.1.2.4. Detectores

Los detectores que se usan son semiconductores⁶, los cuales pueden ser fotoconductores (PbS, PbSe, InAs, InSb) o fotodiodos (InGaAs, Si, Ge). En el caso del InGaAs es necesario que vaya acompañado de un refrigerador termoeléctrico para asegurar el buen funcionamiento.

2.1.3. Parámetros que afectan a la señal

La señal del NIR no varía únicamente por diferencias químicas, sino que se puede dar por muchas otras fuentes de variabilidad¹², algunas de las cuales se describen en la Tabla 2.4. Al desarrollar un método analítico se deberán tener en cuenta para que no se vea comprometida su robustez.

Tabla 2.4 Parámetros que pueden afectar al espectro NIR.

Fuente	Parámetros
Instrumental	Cambio de fungibles
	Longitud de la fibra óptica (si tiene)
Operacional	Temperatura y humedad ambiental
	Posición de la muestra
Muestra	Tamaño de partícula
	Temperatura de la muestra
	Contenido en agua
	Heterogeneidad
	Estado sólido
	Grosor de la muestra (modo transmitancia y transflectancia)

2.1.4. Aplicación NIR en la industria farmacéutica

La espectroscopia NIR es la técnica por excelencia para aplicar sistemas PAT en la industria farmacéutica. Esto es debido a la rapidez con la que se obtiene información de las muestras, tanto química como física, de manera no destructiva. También se suma la versatilidad de instrumentos disponibles en el mercado, que permiten el uso de la técnica en modo *of-*, *at-*, *in-* u *on-line*. La importancia del NIR en la industria farmacéutica se hace evidente tanto a nivel académico como a nivel regulatorio.

Del total de publicaciones encontradas en la base de datos *Web of Science* al buscar los conceptos NIR junto a farmacia y/o PAT se obtiene el gráfico de la Figura 2.3. El número de artículos ha aumentado hasta el año 2015 manteniéndose constante hasta la actualidad.

A nivel regulatorio, en Enero del 2014 la EMA publicó la guía definitiva del uso del NIR y los requisitos al desarrollar un método analítico para la solicitud de registro¹³. La agencia FDA, tiene un guía que va en el mismo sentido desde el 2015, pero aún está en borrador¹⁴. Por otro lado, hay un capítulo dedicado al NIR tanto en la farmacopea europea como en estadounidense (EP 2.2.40. y USP <1119>)^{12,15}.

El uso del NIR en el entorno farmacéutico es muy variado. Así, se puede usar para la identificación de materia prima o productos granel¹⁶⁻¹⁸. También para la cuantificación del contenido en API¹⁹⁻²⁸ para determinar la riqueza o la uniformidad de contenido, y el contenido en agua²⁹⁻³¹. Es una técnica útil para la determinación de propiedades físicas³² o para determinar la cantidad de polimorfo³³ presente en la muestra. Finalmente, el NIR es una técnica ampliamente usada para el seguimiento en tiempo real de distintas operaciones unitarias, como son la granulación^{34,35}, procesos de mezcla³⁶⁻⁴¹ o de recubrimiento de comprimidos⁴².

2.2. ESPECTROSCOPIA RAMAN

La espectroscopia Raman se basa en la dispersión inelástica de la radiación cuando una radiación monocromática interactúa con una molécula, permitiendo el estudio de rotaciones y vibraciones moleculares. El efecto Raman fue observado experimentalmente por Raman y Krishnan en el año 1928⁴³.

2.2.1. Fundamentos

La espectroscopia Raman permite obtener información de la estructura y propiedad de moléculas a partir de las transiciones vibracionales⁴³, pero en vez de basarse en la absorción de la radiación, como en el caso del NIR, se basa en la dispersión de la misma. El fotón de una radiación monocromática interactúa con la molécula produciendo una distorsión en la nube de electrones y creando un estado excitado virtual. La característica de este nuevo estado es que tiene un tiempo de vida muy corto y al relajarse, el fotón es re-emitido, pudiéndose producir una de las situaciones que se esquematizan en la Figura 2.4. La primera situación es cuando el fotón vuelve al estado fundamental, produciéndose una dispersión elástica conocida como dispersión Rayleigh. La frecuencia emitida es la misma que la de la radiación incidente. Esta es la más intensa y a la vez irrelevante ya que no contiene información de la molécula. El segundo escenario se genera cuando

INTRODUCCIÓN

el fotón incidente aparte de una polarización de la nube de electrones, provoca un movimiento nuclear lo que produce una transferencia de energía entre el fotón incidente y la molécula⁴³. En este caso se produce una dispersión inelástica, lo que se conoce como dispersión Raman. Dentro de esta dispersión, se encuentran las líneas Stokes que ocurren cuando la molécula gana un estado vibracional, por lo que la frecuencia del fotón emitido es menor al de la luz incidente. La diferencia de energías corresponde a la energía vibracional de absorción MIR². Por otro lado, cuando la molécula parte de un estado vibracional excitado y pierde un estado vibracional al volver al estado fundamental, entonces la frecuencia del fotón emitido es mayor a la de la luz incidente. Este último es de muy baja intensidad debido a la ley de Boltzmann, según la cual a temperatura ambiente, la mayor parte de las moléculas se encuentra en el estado de menor energía¹. Este último caso solo es interesante estudiarlo cuando la muestra se encuentra a temperatura elevada⁴³.

La dispersión Raman es un proceso débil debido a que solo un fotón se dispersa inelásticamente de un total de 10^6 - 10^8 dispersados, pero el uso de láseres de alta intensidad permite una sensibilidad razonable de la técnica. Se debe encontrar un compromiso entre tener una buena sensibilidad y evitar que se produzca una degradación de la muestra o un proceso de fluorescencia⁴³.

La técnica Raman, como en el caso del NIR, se basa en obtener información de la molécula a partir de sus vibraciones, pero hay diferencias en los tipos de enlaces que pueden ser activos en una y otra, por lo que se trata de técnicas complementarias³. Dado que la condición de que haya actividad en Raman es que se produzca un cambio en la polarizabilidad del enlace durante la vibración, éste es sensible a los enlaces apolares, a los grupos funcionales y vibraciones simétricas. En NIR, la condición es que haya un cambio en el momento dipolar del enlace durante la vibración además de una gran anarmonicidad entre los átomos implicados en los enlaces, por lo que es sensible a enlaces polares, a grupos funcionales y vibraciones asimétricas^{3,44}. A modo de ejemplo, una de las moléculas más activas en NIR, el agua, es totalmente transparente para el Raman.

Cuando la evaluación que se realiza de la señal Raman es cuantitativa, se debe tener en cuenta que la intensidad de la señal es proporcional a la concentración de analito presente en la muestra².

Debido a las ventajas que presentan las técnicas basadas en dispersión Raman (Tabla 2.5), es una técnica atractiva para su uso en la industria farmacéutica como herramienta PAT.

Tabla 2.5 Ventajas de la técnica espectroscópica Raman.

Es una técnica no invasiva ni destructiva, por lo que permite análisis <i>in-</i> y <i>on-line</i>
Registro de espectros rápido, lo que permite el análisis en tiempo real
Alta selectividad
Versátil, permite registrar el espectro de muestras sólidas, líquidas y gaseosas
Permite adquirir espectros a través de contenedores (vidrio o plástico)
Ningún pretratamiento de la muestra o mínimo, por lo que además de contribuir a poder realizar análisis <i>in-</i> y <i>on-line</i> , hay una contribución al medioambiente
El agua es transparente
Posibilidad de uso de fibras ópticas (>100m)
Coste de análisis relativamente bajo

Hay dos inconvenientes que deben tenerse en cuenta a la hora de implementar la técnica. El primero es la posibilidad de que la muestra presente fluorescencia, la cual puede enmascarar completamente la señal Raman. El segundo es la posible degradación térmica de la misma según las condiciones de irradiación y las propiedades de la muestra.

2.2.2. Instrumentación

Un instrumento Raman consta de un láser, ya que la muestra se debe irradiar con una fuente monocromática intensa; el compartimento de la muestra o interfase entre el instrumento y la muestra; un sistema que permita la separación de la radiación en longitudes de ondas discretas y un detector. A continuación se detallan cada una de las partes.

2.2.2.1. Láser

Es necesario el uso de una luz monocromática de alta intensidad ya que la probabilidad que ocurra la dispersión inelástica es baja, por lo que se emplea un láser como fuente de excitación⁴⁵. La intensidad de la dispersión, tal como se muestra en la ecuación Ec 2.1, además de depender la polarizabilidad de la molécula (α), también depende de la cuarta potencia de la frecuencia del láser de excitación (ν_s), y su intensidad (I_0)³. Así, cuanto mayor es la frecuencia del láser, más intensidad de señal se obtiene.

$$I = \frac{8\pi}{9c^4} \nu_s^4 I_0 \alpha^2 \tag{Ec 2.1}$$

Ahora bien, al aumentar la frecuencia del láser, la probabilidad que se produzca un proceso de fluorescencia también crece. Esto es debido a que el estado virtual que se genera al irradiar la molécula, acaba coincidiendo con el estado electrónico excitado. Además, si se aumenta la intensidad del láser, también se puede dar una degradación térmica de la muestra. Se debe encontrar un compromiso entre la sensibilidad y la probabilidad de ocurrencia de fluorescencia y degradación.

En el mercado hay instrumentos con láseres en el rango del visible y NIR (Tabla 2.6). Se escogerá de una longitud de onda concreta dependiendo de las características de las muestras que se van a analizar. El láser a 785nm tiene un buen compromiso entre eficiencia de dispersión y fluorescencia.

Tabla 2.6 Características de los láseres disponibles para instrumentos Raman.

Láser	Características
Visible	+ Aumenta sensibilidad - Aumenta probabilidad fluorescencia + Aumenta probabilidad degradación térmica muestra de muestras con color (posible absorción fotón Raman) - Se acoplan a detectores más baratos
NIR	- Disminuye sensibilidad + Disminuye probabilidad de fluorescencia + Disminuye probabilidad de degradación térmica de las muestras con color

2.2.2.2. Modos de adquisición

En aplicaciones PAT hay principalmente tres modos de adquisición distintas, la retrodispersión (*Backscattering*), el desplazamiento espacial (SORS del inglés *Spatially Offset Raman*) y por transmitancia. Se escogerá un tipo u otro de tecnología dependiendo de la aplicación y de los equipos disponibles. En todos los casos es posible el uso de fibras ópticas.

El modo convencional de adquisición de muestras es el *backscattering*. Este consiste en la obtención de la señal Raman en el mismo lado y en el mismo punto de la radiación incidente. Con este tipo de adquisición se obtiene información de las capas superiores de la muestra. Por lo que puede haber problemas de representatividad si el producto es heterogéneo. Para solventar dicho problema, se suele usar un sistema rotativo que permite aumentar el área muestreada.

Para aumentar la representatividad de la muestra, hay el modo de adquisición SORS con el cual se recoge la señal Raman por el mismo lado que la radiación incidente pero a un punto distinto al del haz incidente, obteniéndose información de capas más internas de la muestra. Esta tecnología es relativamente nueva, por lo que no hay muchos instrumentos que la incorporen.

Finalmente, con el modo de adquisición en transmitancia se detecta la señal Raman en el lado opuesto de la radiación incidente, obteniéndose información de toda la muestra.

2.2.2.3. Sistema de selección de longitud de onda

La función del espectrómetro es separar en distintas longitudes de onda -convertir en valores discretos- la radiación policromática dispersada inelásticamente que sale de la muestra. En el caso de los equipos Raman se encuentran dos tipos, los dispersivos y los FT-Raman. El primero separa las longitudes de onda espacialmente con una rendija. El segundo modula cada longitud de onda, obteniendo una frecuencia modular característica de cada una de ellas, mediante el uso de un interferómetro⁴⁶. Según el tipo de espectrómetro, se usa un láser y detector determinado, los cuales se detallan en la Tabla 2.7 junto con las ventajas y desventajas de cada uno de ellos.

Tabla 2.7 Configuración de equipos Raman según el tipo de espectrómetro usado, la longitud de onda de láser más habitual en mercado y características de cada tipo de equipo.

Equipo	Láser	Detector	Propiedades
Dispersivo	Visible (780nm)	CCD	+ Alta relación S/N + Sensible + Equipo robusto - Resolución espectral intermedia - La resolución varía a lo largo del espectro
FT-Raman	NIR (1064nm)	InGaAs	+ Alta precisión de longitud de onda + Resolución constante a lo largo del espectro - Baja relación S/N

2.2.2.4. Detector

Como se muestra en la Tabla 2.7, hay dos tipos de detectores según el equipo que se utiliza. El dispositivos de carga acoplada (CCD del inglés *Charge Coupled Devices*) y fotodiodos, principalmente InGaAs.

En el caso de CCD, los espectros se registran mediante la generación huecos-electrones que se producen en un material semiconductor fotosensible, generados por fotones con suficiente energía. El CCD necesita que se enfría a -30°C para reducir el ruido de corriente de oscuridad (generación aleatoria de hueco y electrones), fenómeno que aumenta con la temperatura⁴⁵. Para que la sensibilidad sea alta, la temperatura del detector debe disminuir hasta que la intensidad de corriente oscura sea satisfactoriamente baja.

2.2.3. Parámetros que afectan a la señal

El espectro Raman, como en el caso del NIR, se puede ver afectado por parámetros ajenos a la característica con la que se está interesado. Algunos de los más importantes se describen en la Tabla 2.8.

Tabla 2.8 Parámetros que pueden afectar al espectro Raman

Fuente	Parámetros
Instrumental	Ruido de disparo (<i>shot</i>): es el resultado de la naturaleza aleatoria de la luz. Es una fuente de ruido inevitable
	Ruido térmico: en el caso de los detectores CCD, pero que cuando está enfriado este es despreciable
	Rayos cósmicos (partículas de alta energía), estos se concentran en uno o dos elementos (píxeles) del detector CCD, apareciendo un pico muy estrecho y de gran intensidad en el espectro
	Intensidad del láser
	Longitud fibra óptica (si tiene)
Operacional	Temperatura y humedad ambiental
	Posición de la muestra
Muestra	Fluorescencia
	Color de la muestra
	Heterogeneidad
	Estado sólido
	Grueso de la muestra (modo transmitancia)

2.2.4. Aplicación Raman en la industria farmacéutica

Por sus características, la espectroscopia Raman es otra de las técnicas que cumplen con las características para ser usadas como herramienta PAT. Regulatoriamente, no hay ninguna guía específica para la implementación en el entorno farmacéutico, pero dado que el desarrollo de un método es parecido al NIR, se pueden usar las guías publicadas por la EMA y la FDA para esta técnica⁴⁷. En la farmacopea europea y estadounidense hay un capítulo específico dedicado a la técnica Raman (EP 2.2.48. y USP <1120>)^{44,48}.

Las publicaciones relacionadas con el desarrollo de métodos Raman y su implementación en el campo farmacéutico han aumentado anualmente hasta la actualidad (Figura 2.5). Esto se debe, en

parte, a la mejora instrumental. Entre sus aplicaciones se encuentran la identificación de materia prima^{49,50} y la detección de falsificaciones⁵¹⁻⁵⁵; cuantificación de API tanto para determinar la riqueza como la uniformidad de contenido⁵⁶⁻⁶¹; la cuantificación de polimorfismo⁶²⁻⁶⁵ o el seguimiento de operaciones unitarias^{63,65-71}.

2.3. ESPECTROMETRÍA DE MOVILIDAD IÓNICA

La espectrometría de movilidad iónica (IMS del inglés *Ion Mobility Spectrometry*) es una técnica de separación, basada en la migración de iones en fase gas que atraviesan un campo eléctrico homogéneo a presión atmosférica y a contracorriente de un gas inerte. Las moléculas son separadas según su movilidad iónica; característica que depende de la relación masa/carga, de su forma y su tamaño. Los principios de la movilidad de los iones en un campo eléctrico ya se conocían a principios del siglo XX, aunque no fue hasta los años 70 cuando se desarrolló la instrumentación.

2.3.1. Fundamentos de IMS

La separación en IMS se basa en la diferencia de velocidad de las especies cargadas bajo la acción de un campo eléctrico, que a su vez depende de su relación masa-carga y /o su estructura. La velocidad de deriva de la nube de iones de una especie química (v_d) es proporcional al campo eléctrico aplicado (E), según la ecuación Ec 2.2.

$$v_d = K \cdot E \quad \text{Ec 2.2}$$

La constante de proporcionalidad es lo que se define como constante de movilidad (K), la cual se describe a nivel molecular por la ecuación de Mason-Schamp, ecuación Ec 2.3.

$$K = \left(\frac{3ez}{16N}\right) \left(\sqrt{\frac{2\pi}{kT}}\right) \left(\sqrt{\frac{m+M}{mM}}\right) \left(\frac{1+\alpha}{\Omega}\right) \quad \text{Ec 2.3}$$

donde e es la carga del electrón (1.602×10^{-19} C), z es la carga del ión, N es la densidad del gas de deriva, k la constante de Boltzmann ($1,381 \times 10^{-23}$ J K⁻¹) y T la temperatura en Kelvin, m y M son la masa del ión y del gas de deriva respectivamente, α es un término de corrección dependiente de la masa y Ω es la sección transversal de colisión, derivada de las integraciones de todos los posibles ángulos de dispersión de la sección transversal de la especie ión-molécula neutra de gas de deriva⁷²⁻⁷⁴.

Como la constante de movilidad (K) es función de la densidad de gas de deriva y por tanto presenta una alta dependencia de la temperatura y de la presión, usualmente se normaliza a condiciones estándar (273K y 760Torr) mediante la ecuación Ec 2.4. Esta constante normalizada, conocida

como movilidad reducida (K_0), es necesaria para comparar datos experimentales que se obtienen bajo diferentes condiciones experimentales.

$$K_0 = K \left(\frac{273}{T} \right) \left(\frac{P}{760} \right) \quad \text{Ec 2.4}$$

2.3.2. Instrumentación IMS

Un instrumento IMS consta principalmente de dos partes, la región de ionización y la de deriva. En la primera es donde la sustancia química se ioniza, mientras que en la segunda es donde se produce la migración de los iones generados. En la Figura 2.6 se presenta un esquema más detallado de los componentes de instrumento IMS.

A continuación se detalla cada uno de ellos, así como los procesos que tienen lugar en él.

2.3.2.1. Sistema de introducción de muestra

El objetivo del sistema de introducción de muestra es suministrar la muestra vaporizada a la región de ionización con la mínima pérdida de muestra y sin distorsionar su información química. El sistema usado depende del tipo de muestra de la cual se disponga⁷⁴:

- Las muestras gaseosas se pueden introducir directamente a la región de ionización, o bien usar sistemas basados en membranas. También se encuentran sistemas de diluciones y sistemas de pre-concentración.
- Para muestras líquidas, los sistemas más usados son membranas semipermeables, introducción por spray o micro-extracción en fase sólida (SPME del inglés *Solid-Phase Micro-Extraction*).
- Para muestras sólidas se usan sistemas de vaporización térmica, vaporización por calentamiento mediante láser o ablación.

2.3.2.2. Fuente de ionización

La fuente de ionización sirve para ionizar las moléculas neutras de la muestra a presión ambiental⁷³. Las más empleadas son las radioactivas de ⁶³Ni, ionización por *electrospray* (ESI), corona de descarga, fotoionizador, MALDI y flama⁷⁴. Dado que la fuente usada para el desarrollo de esta tesis ha sido la de ⁶³Ni, a continuación se explica el proceso que tiene lugar.

La ionización tiene lugar a presión atmosférica y consiste en dos etapas. Primero se forman los iones reactantes mediante la emisión de radiación β del ⁶³Ni. La colisión de los electrones de alta energía con el nitrógeno y el oxígeno del aire genera una reacción en cadena con las que se acaban obteniendo los iones reactantes. Este proceso puede tener lugar en polaridad positiva (ión reactante $H^+(H_2O)_n$) o negativa (ión reactante $(H_2O)_nO^{2-}$). El nivel de humedad del aire y la temperatura determinará el valor de n ⁷³. La segunda etapa es la ionización de la muestra al producirse una transferencia de carga con el ión reactante. A continuación se muestran las reacciones que tienen lugar tanto en modo positivo como en negativo:

- Modo positivo: esta reacción es favorable si la afinidad de la molécula con el ión reactante es mayor que con el agua (moléculas con afinidad protónica mayor que el agua).
- Modo negativo: favorecida cuando las moléculas presentan electroafinidad.

En ambos casos se puede usar un gas reactivo, conocido como dopante, para reducir las interferencias de la matriz y así aumentar la selectividad⁷³. Si se trabaja en modo positivo el gas reactivo controla la transferencia de protones. Se emplea acetona, amoníaco, DMSO, nicotinamida o nonilamina. En polaridad negativa el gas reactivo controla las reacciones de acoplamiento aniónico, y los más empleados son pequeños clorocarbonos primarios, como el hexacloroetano, para producir ion cloruro⁷⁴.

Por último, una desventaja de la ionización con ⁶³Ni respecto a otros modos de ionización, es que si hay más de una especie con las características apropiadas para poderse ionizar, hay una competitividad entre ellas durante dicho proceso. Esto es debido a que el reactivo ionizante es limitante, por lo que la información analítica que se obtiene es sesgada.

2.3.2.3. Tubo de deriva

Una vez ionizada la muestra, esta se introduce al tubo de deriva mediante una rejilla obturadora. El flujo de iones se mueve bajo la influencia de un campo eléctrico y contra un gas de deriva. El

fenómeno que tiene lugar se ha explicado en el apartado 2.3.1. *Fundamentos de IMS*. Los iones se separan según su masa, carga iónica y forma. Se suele usar un calibrante interno -nicotinamida en modo positivo y 4-nitrobenzonitrilo en modo negativo- por lo que el coeficiente de movilidad reducida se calcula con la ecuación Ec 2.5.

$$K_0^{an} = K_0^{cal} \frac{t_d^{cal}}{t_d^{an}} \quad \text{Ec 2.5}$$

2.3.2.4. Detector

Se usa un plato de Faraday como detector⁷⁴, de modo que los iones se neutralizan colisionando con él y se produce un flujo de corriente que es amplificada y convertida en voltaje. El tiempo que tarda un ión desde la rejilla obturadora al detector es el tiempo de deriva.

Se obtiene un espectro de movilidad que consiste en la respuesta del detector frente al tiempo de deriva (ms). Con este se puede obtener información cualitativa con la K_0 y cuantitativa con la amplitud de los picos.

2.3.3. **Aplicación de IMS en la industria farmacéutica**

Las ventajas que presenta la técnica IMS son su alta sensibilidad y rapidez en la adquisición de datos. También hay disponibles equipos que permiten el análisis in-situ^{73,75}. Además, el volumen de disolvente necesario para la preparación de la muestra es muy bajo y hay muy poca generación de residuos, los que convierte el IMS en una técnica de bajo coste y respetuoso con el medio ambiente. La mayor desventaja es su limitada resolución⁷⁶. Otra desventaja a tener en cuenta es que la generación de iones reactantes es limitada, por lo que si la muestra tiene más de un analito habrá competencia para ionizarse, pudiéndose producir pérdidas de información.

Históricamente el IMS ha sido empleado en el campo de la seguridad y militar para detectar componentes orgánicos volátiles y semi-volátiles, tales como trazas de explosivos, de agentes químicos de guerra y de drogas ilegales^{73,77,78}. Sin embargo, gracias a sus características, esta técnica ha ganado importancia en los últimos años en otros campos⁷⁹, como son el alimentario⁸⁰, el biomédico⁸¹ o el ambiental⁸². Las publicaciones existentes en lo referente al campo farmacéutico, en el gráfico de la Figura 2.7, no son muy altas. Esto puede ser debido a la falta de casas comerciales que ofrezcan un instrumento con las prestaciones necesarias para ser implementadas en el entorno regulado. Pero las aplicaciones publicadas referentes a la validación y/o verificación de limpieza⁸³⁻⁸⁷ son de elevado interés, sobre todo por la sensibilidad y rapidez que ofrece la técnica.

2.4. DATOS DE PROCESO

Se entiende como datos de proceso todos aquellos datos que se generan durante la fabricación de un lote de medicamento. Estos incluyen los parámetros de calidad de materias primas, los parámetros de proceso y los atributos de calidad del producto en granel y terminado. También todos los datos derivados de la fabricación, como son, por ejemplo, datos de los procedimientos de limpieza.

Un lote de producto terminado consta de un número determinado de operaciones unitarias. Cada operación unitaria tiene asociada una orden de trabajo. Cada producto intermedio tiene un lote específico y único. Además, cada materia prima usada puede provenir de un solo lote o de varios distintos. Así, para un solo lote de producto terminado que se encuentra en el mercado se obtiene una gran cantidad de datos.

2.4.1. Parámetros de calidad de materias primas

A cada material de partida y material de acondicionamiento se le realizan las pruebas necesarias para aceptar la validez de un lote. Todos los resultados de las pruebas realizadas se consideran parámetros de calidad de materia prima y se registran y almacenan en el sistema de gestión de información de laboratorio (LIMS del inglés *Laboratory Information Management System*).

2.4.2. Parámetros de proceso

Para cada una de las operaciones unitarias implicadas en la fabricación de un producto se registran los parámetros de proceso implicados en una base de datos. Estos se pueden usar para entender la influencia que tienen individualmente y/o conjuntamente en la calidad del producto terminado.

Las características de este tipo de datos es que el volumen generado es muy alto. Además, dado que el tiempo de fabricación no es exactamente el mismo para cada lote de fabricación, la longitud de

vectores puede variar lote a lote. Esto hace que sea necesario realizar un tratamiento previo al análisis.

2.4.3. Atributos de calidad

Son los parámetros de calidad evaluados en productos intermedios, en el producto granel o en el producto terminado. Todos ellos deben encontrarse dentro de las especificaciones declaradas para que puedan liberarse en el mercado. Estos parámetros pueden ser cualitativos o cuantitativos. Todos ellos, como en el caso de los atributos de los materiales de partida, se encuentran almacenados en el LIMS.

2.4.4. Aplicaciones en la industria farmacéutica

La gran cantidad de datos almacenados contienen una información muy valiosa para poder conocer mucho mejor el proceso de fabricación y poder entender las posibles relaciones entre los CMA, CPP y CQA. Por lo que con esto se podrá, entre otros:

- Optimizar procesos que están en curso.
- Detectar causas raíces en lotes fuera de especificaciones, que permanecen latentes si no se entienden las relaciones.
- Sentar las bases para realizar una propuesta de desarrollo por QbD, siguiendo las directrices de la ICH Q8, Q9, Q10, Q11 y Q12. Cuando un proceso es nuevo, es inevitable plantear todo los diseños de experimentos necesarios para comprender lo que se tiene entre manos. Pero cuando el proceso ya está implementado con el enfoque tradicional, el análisis retrospectivo de datos y las propias experiencias simplificarán el número de experimentos a realizar.

2.5. BIBLIOGRAFÍA

1. Burns, D. A. & Ciurczak, E. W. *Handbook of Near Infrared Analysis*. (CRC Press Taylor and Francis, 2008).
2. Siesler, H. W., Ozaki, Y., Kawata, S. & Heise, H. M. *Near-infrared spectroscopy. Principles, instruments, applications*. (Wiley-VCH, 2002).
3. Ríos Castro, Á., Moreno Bondi, M. C. & Simonet Suau, B. M. *Técnicas espectroscópicas en química analítica Volumen I*. (Editorial Síntesis, 2012).
4. Reich, G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug Deliv. Rev.* **57**, 1109–1143 (2005).
5. Pasquini, C. Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Anal. Chim. Acta* **1026**, 8–36 (2018).
6. Simpson, M. B. Near-infrared Spectroscopy for Process Analytical Technology: Theory , Technology and Implementation. in *Process Analytical Technology* (ed. Bakeev, K. A.) 107–155 (John Wiley & Sons Ltd., 2010).

7. Sablinskas, V. Instrumentation. in *Handbook of Spectroscopy* (eds. Gauglitz, G. & Moore, D. S.) 39–70 (Wiley-VCH, 2014).
8. Alcalà, M. *et al.* Near-infrared Spectroscopy in Laboratory and Process Analysis. in *Encyclopedia of Analytical Chemistry* (eds. Meyers, R. A. & Meyers, R. A.) (2012).
9. Agelet, L. E. & Hurburgh, C. R. A Tutorial on Near Infrared Spectroscopy and Its Calibration. *Crit. Rev. Anal. Chem.* **40**, 246–260 (2010).
10. Alcalà Bernàrdez, M. Utilización de la Espectroscopia NIR en el Control Analítico de la Industria Farmacéutica Desarrollos iniciales en PAT. (2006).
11. Ciurczak, E. W. & Igne, B. *Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy*. CRC Press (2002).
12. European Pharmacopoeia. Ph. Eur. 9.0 Chapter 2.2.40. Near-infrared spectroscopy. (2017).
13. EMA. Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations. *Eur. Med. Agency* **44**, 1–28 (2014).
14. FDA. Development and Submission of Near Infrared Analytical Procedures Guidance for Industry. 24 (2015).
15. United States Pharmacopeia. USP 40–NF 35 <1119> Near Infrared Spectroscopy. (2016).
16. Gemperline, P. J., Webber, L. D. & Cox, F. O. Raw Materials Testing Using Soft Independent Modeling of Class Analogy Analysis of Near-Infrared Reflectance Spectra. *Anal. Chem.* **61**, 138–144 (1989).
17. Kemper, M. S. & Luchetta, L. M. A guide to raw material analysis using near infrared spectroscopy. *J. Near Infrared Spectrosc.* **11**, 155–174 (2003).
18. Sun, L. *et al.* Pharmaceutical Raw Material Identification Using Miniature Near-Infrared (MicroNIR) Spectroscopy and Supervised Pattern Recognition Using Support Vector Machine. *Appl. Spectrosc.* **70**, 816–825 (2016).
19. Chavez, P. F. *et al.* Active content determination of pharmaceutical tablets using near infrared spectroscopy as Process Analytical Technology tool. *Talanta* **144**, 1352–1359 (2015).
20. Eustaquio, A., Graham, P., Jee, R. D., Moffat, A. C. & Trafford, A. D. Quantification of paracetamol in intact tablets using near-infrared transmittance spectroscopy. *Analyst* **123**, 2303–2306 (1998).
21. Ito, M. *et al.* Development of a method for the determination of caffeine anhydrate in various designed intact tables by near-infrared spectroscopy: A comparison between reflectance and transmittance technique. *J. Pharm. Biomed. Anal.* **47**, 819–827 (2008).
22. Chalus, P., Roggo, Y., Walter, S. & Ulmschneider, M. Near-infrared determination of active substance content in intact low-dosage tablets. *Talanta* **66**, 1294–1302 (2005).
23. Sulub, Y., LoBrutto, R., Vivilecchia, R. & Wabuye, B. Near-infrared multivariate calibration updating using placebo: A content uniformity determination of pharmaceutical tablets. *Vib. Spectrosc.* **46**, 128–134 (2008).
24. Sulub, Y., LoBrutto, R., Vivilecchia, R. & Wabuye, B. W. Content uniformity determination of pharmaceutical tablets using five near-infrared reflectance spectrometers: A process analytical technology (PAT) approach using robust multivariate calibration transfer algorithms. *Anal. Chim. Acta* **611**, 143–150 (2008).
25. Xiang, D. *et al.* Evaluation of transmission and reflection modalities for measuring content uniformity of pharmaceutical tablets with near-infrared spectroscopy. *Appl. Spectrosc.* **63**, 33–47 (2009).

26. Trafford, A. D., Jee, R. D., Moffat, A. C. & Graham, P. A rapid quantitative assay of intact paracetamol tablets by reflectance near-infrared spectroscopy. *Analyst* **124**, 163–167 (1999).
27. Tomuta, I., Rus, L., Iovanov, R. & Rus, L. L. High-throughput NIR-chemometric methods for determination of drug content and pharmaceutical properties of indapamide tablets. *J. Pharm. Biomed. Anal.* **84**, 285–292 (2013).
28. Bu, D., Wan, B. & McGeorge, G. A discussion on the use of prediction uncertainty estimation of NIR data in partial least squares for quantitative pharmaceutical tablet assay methods. *Chemom. Intell. Lab. Syst.* **120**, 84–91 (2013).
29. Clavaud, M. *et al.* Global regression model for moisture content determination using near-infrared spectroscopy. *Eur. J. Pharm. Biopharm.* **119**, 343–352 (2017).
30. Clavaud, M. *et al.* Moisture content determination in an antibody-drug conjugate freeze-dried medicine by near-infrared spectroscopy: A case study for release testing. *J. Pharm. Biomed. Anal.* **131**, 380–390 (2016).
31. Mainali, D., Li, J., Yehl, P. & Chetwyn, N. Development of a comprehensive near infrared spectroscopy calibration model for rapid measurements of moisture content in multiple pharmaceutical products. *J. Pharm. Biomed. Anal.* **95**, 169–175 (2014).
32. Blanco, M., Alcalá, M., González, J. M. & Torras, E. A Process Analytical Technology Approach Based on Near Infrared Spectroscopy: Tablet Hardness, Content Uniformity, and Dissolution Test Measurements of Intact Tablets. *J. Pharm. Sci.* **95**, 2137–2144 (2006).
33. da Silva, V. H., da Silva, J. J. & Pereira, C. F. Portable near-infrared instruments: Application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer. *J. Pharm. Biomed. Anal.* **134**, 287–294 (2017).
34. Peinado, A., Hammond, J. & Scott, A. Development, validation and transfer of a Near Infrared method to determine in-line the end point of a fluidised drying process for commercial production batches of an approved oral solid dose pharmaceutical product. *J. Pharm. Biomed. Anal.* **54**, 13–20 (2011).
35. Alcalá, M., Blanco, M., Bautista, M. & González, J. M. On-Line Monitoring of A Granulation Process By NIR Spectroscopy. *J. Pharm. Sci.* **99**, 336–345 (2010).
36. Martínez, L., Peinado, A., Liesum, L. & Betz, G. Use of near-infrared spectroscopy to quantify drug content on a continuous blending process: Influence of mass flow and rotation speed variations. *Eur. J. Pharm. Biopharm.* **84**, 606–615 (2013).
37. Martínez, L., Peinado, A. & Liesum, L. In-line quantification of two active ingredients in a batch blending process by near-infrared spectroscopy: Influence of physical presentation of the sample. *Int. J. Pharm.* **451**, 67–75 (2013).
38. Sulub, Y., Königsberger, M. & Cheney, J. Blend uniformity end-point determination using near-infrared spectroscopy and multivariate calibration. *J. Pharm. Biomed. Anal.* **55**, 429–434 (2011).
39. Corredor, C. C. *et al.* Analytical Method Quality by Design for an On-Line Near-Infrared Method to Monitor Blend Potency and Uniformity. *J. Pharm. Innov.* **10**, 47–55 (2014).
40. Ward, H. W., Blackwood, D. O., Polizzi, M. & Clarke, H. Monitoring blend potency in a tablet press feed frame using near infrared spectroscopy. *J. Pharm. Biomed. Anal.* **80**, 18–23 (2013).
41. Sulub, Y. *et al.* Real-time on-line blend uniformity monitoring using near-infrared reflectance spectrometry: A noninvasive off-line calibration approach. *J. Pharm. Biomed. Anal.* **49**, 48–54 (2009).
42. Römer, M., Heinämäki, J., Strachan, C., Sandler, N. & Yliruusi, J. Prediction of Tablet Film-coating Thickness Using a Rotating Plate Coating System and NIR Spectroscopy. *AAPS PharmSciTech* **9**,

- 1047–1053 (2008).
43. Smith, EwenDent, G. *Modern Raman spectroscopy-A practical approach*. (Wiley, 2005).
 44. European Pharmacopoeia. Ph. Eur. 9.0 Chapter 2.2.48. Raman spectroscopy. (2017).
 45. Lewis, I. R. & Edwards, H. G. M. *Handbook of Raman Spectroscopy*. (Taylor & Francis, 2001).
 46. McCreery, R. L. *Raman Spectroscopy for Chemical Analysis. Book 8*, (John Wiley & Sons, INC., 2000).
 47. Andrews, D. *et al.* Analytical Method Development Using Transmission Raman Spectroscopy for Pharmaceutical Assays and Compliance with Regulatory Guidelines—Part I: Transmission Raman Spectroscopy and Method Development. *J. Pharm. Innov.* **13**, 121–132 (2018).
 48. United States Pharmacopeia. USP 40–NF 35 <1120> Raman spectroscopy.
 49. Bloomfield, M. *et al.* Non-invasive identification of incoming raw pharmaceutical materials using Spatially Offset Raman Spectroscopy. *J. Pharm. Biomed. Anal.* **76**, 65–69 (2013).
 50. Gryniewicz-Ruzicka, C. M., Rodriguez, J. D., Arzhantsev, S., Buhse, L. F. & Kauffman, J. F. Libraries, classifiers, and quantifiers: A comparison of chemometric methods for the analysis of Raman spectra of contaminated pharmaceutical materials. *J. Pharm. Biomed. Anal.* **61**, 191–198 (2012).
 51. Fraser, S. J. *et al.* Simultaneous qualitative and quantitative analysis of counterfeit and unregistered medicines using Raman spectroscopy. *J. Raman Spectrosc.* **44**, 1172–1180 (2013).
 52. Dégardin, K., Roggo, Y., Been, F. & Margot, P. Detection and chemical profiling of medicine counterfeits by Raman spectroscopy and chemometrics. *Anal. Chim. Acta* **705**, 334–341 (2011).
 53. Neuberger, S. & Neusüß, C. Determination of counterfeit medicines by Raman spectroscopy: Systematic study based on a large set of model tablets. *J. Pharm. Biomed. Anal.* **112**, 70–78 (2015).
 54. Lanzarotta, A., Lorenz, L., Batson, J. C. S. & Flurer, C. Development and implementation of a pass/fail field-friendly method for detecting sildenafil in suspect pharmaceutical tablets using a handheld Raman spectrometer and silver colloids. *J. Pharm. Biomed. Anal.* **146**, 420–425 (2017).
 55. Dégardin, K., Guillemain, A. & Roggo, Y. Comprehensive Study of a Handheld Raman Spectrometer for the Analysis of Counterfeits of Solid-Dosage Form Medicines. *J. Spectrosc.* **2017**, (2017).
 56. Mazurek, S. & Szostak, R. Quantitative determination of diclofenac sodium and aminophylline in injection solutions by FT-Raman spectroscopy. *J. Pharm. Biomed. Anal.* **40**, 1235–1242 (2006).
 57. Braun, D. E. *et al.* Simultaneous quantitative analysis of ternary mixtures of d-mannitol polymorphs by FT-Raman spectroscopy and multivariate calibration models. *Int. J. Pharm.* **385**, 29–36 (2010).
 58. De Bleye, C. *et al.* Determination of 4-aminophenol in a pharmaceutical formulation using surface enhanced Raman scattering: From development to method validation. *Talanta* **116**, 899–905 (2013).
 59. Mazurek, S. & Szostak, R. Quantification of active ingredients in pharmaceutical suspensions by FT Raman spectroscopy. *Vib. Spectrosc.* **93**, 57–64 (2017).
 60. Cahyadi, C., Sasithornwetchakun, W., Topark-Ngarm, A., Chana, L. W. & Henga, P. W. S. An evaluation of content uniformity of tablets prepared by quasi-continuous drug coating using near infrared and Raman spectroscopy. *J. Near Infrared Spectrosc.* **22**, 239–248 (2014).
 61. Brouckaert, D., Uyttensprot, J. S., Broeckx, W. & De Beer, T. Development and validation of an at-line fast and non-destructive Raman spectroscopic method for the quantification of multiple components in liquid detergent compositions. *Anal. Chim. Acta* **941**, 26–34 (2016).

62. Farias, M. A. dos S., Soares, F. L. F. & Carneiro, R. L. Crystalline phase transition of ezetimibe in final product, after packing, promoted by the humidity of excipients: Monitoring and quantification by Raman spectroscopy. *J. Pharm. Biomed. Anal.* **121**, 209–214 (2016).
63. Netchacovitch, L. *et al.* Development of an analytical method for crystalline content determination in amorphous solid dispersions produced by hot-melt extrusion using transmission Raman spectroscopy: A feasibility study. *Int. J. Pharm.* **530**, 249–255 (2017).
64. Hennigan, M. C. & Ryder, A. G. Quantitative polymorph contaminant analysis in tablets using Raman and near infra-red spectroscopies. *J. Pharm. Biomed. Anal.* **72**, 163–171 (2013).
65. Van Renterghem, J. *et al.* Elucidation and visualization of solid-state transformation and mixing in a pharmaceutical mini hot melt extrusion process using in-line Raman spectroscopy. *Int. J. Pharm.* **517**, 119–127 (2017).
66. Harting, J. & Kleinebudde, P. Development of an in-line Raman spectroscopic method for continuous API quantification during twin-screw wet granulation. *Eur. J. Pharm. Biopharm.* **125**, 169–181 (2018).
67. Grymonpré, W. *et al.* Downstream processing from hot-melt extrusion towards tablets: A quality by design approach. *Int. J. Pharm.* **531**, 235–245 (2017).
68. Allan, P. *et al.* In situ monitoring of powder blending by non-invasive Raman spectrometry with wide area illumination. *J. Pharm. Biomed. Anal.* **76**, 28–35 (2013).
69. Riolo, D. *et al.* Raman spectroscopy as a PAT for pharmaceutical blending: Advantages and disadvantages. *J. Pharm. Biomed. Anal.* **149**, 329–334 (2018).
70. Knop, K. & Kleinebudde, P. PAT-tools for process control in pharmaceutical film coating applications. *Int. J. Pharm.* **457**, 527–536 (2013).
71. Esmonde-White, K. A., Cuellar, M., Uerpmann, C., Lenain, B. & Lewis, I. R. Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. *Anal. Bioanal. Chem.* **409**, 637–649 (2017).
72. Creaser, C. S. *et al.* Ion mobility spectrometry: A review. Part 1. Structural analysis by mobility measurement. *Analyst* **129**, 984–994 (2004).
73. Borsdorf, H. & Eiceman, G. A. Ion Mobility Spectrometry: Principles and Applications. *Appl. Spectrosc. Rev.* **41**, 323–375 (2006).
74. G.A.Eiceman; Z.Karpas. *Ion mobility spectrometry*. (CRC Press Taylor and Francis, 2005).
75. Baumbach, J. I. Process analysis using ion mobility spectrometry. *Anal. Bioanal. Chem.* **384**, 1059–70 (2006).
76. Hill, H. H. & Simpson, G. Capabilities and limitations of ion mobility spectrometry for field screening applications. *F. Anal. Chem. Technol.* **1**, 119–134 (1997).
77. Asbury, G. R. & Jr, H. H. H. Analysis of explosives using electrospray ionization / ion mobility spectrometry (ESI / IMS). *Talanta* **50**, 1291–1298 (2000).
78. Ewing, R. G., Atkinson, D. A., Eiceman, G. A. & Ewing, G. J. A critical review of ion mobility spectrometry for the detection of explosives and explosive related compounds. *Talanta* **54**, 515–529 (2001).
79. Armenta, S., Alcalá, M. & Blanco, M. A review of recent, unconventional applications of ion mobility spectrometry (IMS). *Anal. Chim. Acta* **703**, 114–23 (2011).
80. Vautz, W. *et al.* Ion mobility spectrometry for food quality and safety. *Food Addit. Contam.* **23**, 1064–73 (2006).

81. Garrido-Delgado, R., Arce, L., Pérez-Marín, C. C. & Valcárcel, M. Use of ion mobility spectroscopy with an ultraviolet ionization source as a vanguard screening system for the detection and determination of acetone in urine as a biomarker for cow and human diseases. *Talanta* **78**, 863–8 (2009).
82. Márquez-Sillero, I., Aguilera-Herrador, E., Cárdenas, S. & Valcárcel, M. Ion-mobility spectrometry for environmental analysis. *TrAC Trends Anal. Chem.* **30**, 677–690 (2011).
83. Strege, M. A., Kozerski, J., Juarbe, N. & Mahoney, P. At-Line Quantitative Ion Mobility Spectrometry for Direct Analysis of Swabs for Pharmaceutical Manufacturing Equipment Cleaning Verification. *Anal. Chem.* **80**, 3040–3044 (2008).
84. Baert, B., Boonen, J., Thierens, C. & de Spiegeleer, B. Ion mobility spectrometry of talarozole, a new azole drug, in cleaning quality control. *Int. J. Ion Mobil. Spectrom.* **14**, 109–116 (2011).
85. Strege, M. A. Total residue analysis of swabs by ion mobility spectrometry. *Anal. Chem.* **81**, 4576–4580 (2009).
86. Armenta, S., Alcalà, M., Blanco, M. & González, J. M. Ion mobility spectrometry for the simultaneous determination of diacetyl midecamycin and detergents in cleaning validation. *J. Pharm. Biomed. Anal.* (2013).
87. Qin, C., Granger, A., Papov, V., McCaffrey, J. & Norwood, D. L. Quantitative determination of residual active pharmaceutical ingredients and intermediates on equipment surfaces by ion mobility spectrometry. *J. Pharm. Biomed. Anal.* (2010).

3. TÉCNICAS DE ANÁLISIS MULTIVARIABLE DE DATOS

Las técnicas de análisis multivariable son el conjunto de herramientas empleadas para estudiar objetos definidos por más de una variable (generalmente un número elevado de ellas) y así poder transformar la información disponible en conocimiento. Cuando estos datos provienen de técnicas o señales analíticas se habla de quimiometría. La quimiometría se puede definir como “una disciplina de la química que utiliza métodos matemáticos, estadísticos y otros con el objetivo de (a) diseñar o seleccionar procedimientos de medida y experimentos óptimo, y (b) obtener el máximo de información química relevante analizando los datos químicos”¹. Esta disciplina incluye diseño de experimentos, exploración de datos y desarrollo de modelos predictivos².

Como se ha comentado, el uso de la quimiometría es imprescindible en el empleo de la espectroscopia NIR y Raman como técnicas analíticas, para poder extraer la información útil (exploración de datos), agrupar muestras según sus características (métodos de clasificación) o relacionar la señal analítica con una o varias propiedades cuantificables de interés (métodos cuantitativos).

Por otro lado, la fabricación de un medicamento proporciona una cantidad elevada de datos, procedentes de los atributos de las materias primas, variables de proceso, atributos de calidad de productos intermedios y/o acabados... por lo que el empleo del análisis multivariable de los datos permite obtener un conocimiento profundo del proceso. Esto es una condición indispensable para avanzar hacia una fabricación basada en el QbD o en la fabricación en continuo.

Las herramientas usadas en esta tesis para el análisis multivariable de datos son los métodos que usualmente se emplean en quimiometría, por lo que de ahora en adelante se van a usar estos conceptos indistintamente, tanto si los datos provienen de señales químicas como si no. En este capítulo se definirán los métodos quimiométricos de forma genérica (diseño de experimentos, pretratamiento matemático de datos, análisis cualitativo y análisis cuantitativo) así como los pasos que se deben seguir para desarrollar un modelo NIR o Raman en el entorno regulado de la industria farmacéutica.

3.1. DISEÑO DE EXPERIMENTOS

El diseño de experimentos (DoE del inglés *Design of Experiments*) es una técnica que permite planificar experimentos relevantes maximizando la obtención de información y minimizando los costes. Se pueden usar para estudiar los efectos de factores que pueden tener en una respuesta determinada y las interacciones entre ellos (*screening*) o para modelar y optimizar procesos³.

INTRODUCCIÓN

La finalidad es obtener un modelo matemático capaz de relacionar la respuesta de interés con las variables experimentales. Este modelo contiene términos lineales (efectos directos de cada factor a la respuesta), términos de interacción entre factores (de primer orden cuando la interacción es entre dos factores, de segundo orden cuando es entre tres factores...) y finalmente, términos cuadráticos (permiten determinar relaciones no lineales entre las variables y la respuesta). Para realizar estudios de *screening*, se usan los dos primeros, efectos directos e interacciones, mientras que para estudios de optimización se introducen los términos cuadráticos⁴.

En todos los casos, el proceso para llevar a cabo un DoE empieza con la definición del objetivo del estudio. Después, es necesario detectar todas las variables (factores) que podrían tener efecto en la respuesta y definir el rango y niveles de cada uno de ellos. A continuación se planifican los experimentos que se llevarán a cabo mediante la elección de un diseño de experimentos, seguido de la realización de los experimentos, de manera aleatoria para evitar errores sistemáticos. Finalmente se analizan los datos para obtener la información⁵.

Hay muchas opciones para escoger el diseño de experimentos adecuado. Esta elección se realizará según las necesidades y objetivos del estudio. En la Tabla 3.1 se describen algunos de los más empleados para realizar un *screening*. Con el factorial completo es con el que se obtiene mayor información. Ahora bien, el número de experimentos a realizar aumenta exponencialmente con la cantidad de factores incluidos en el estudio. Así que, si el número de experimentos debe disminuir, se puede escoger otro DoE, teniendo presente que siempre se sacrifica información a cambio.

Tabla 3.1 Diseño de experimentos para realizar estudios de *screening*.

Diseño de experimentos	Características
Factorial Completo ⁴	<ul style="list-style-type: none">- 2^k experimentos (k es el número de factores y 2 el número de niveles por factor).- Se suelen añadir 3 experimentos en el centro para calcular intervalos de confianza.- Se detectan todos los factores e interacciones entre ellos que tienen efecto en la respuesta.- El número de experimentos aumenta exponencialmente con el número de factores.
Factorial fraccionado ⁴	<ul style="list-style-type: none">- 2^{k-p} experimentos (k es el número de factores, p es la magnitud de la fracción y 2 el número de niveles por factor).- Se realiza una fracción de experimentos del factorial completo.- Al disminuir el número de experimentos se explora parcialmente el dominio experimental.- Aparecen confusiones. Se debe encontrar un compromiso entre el número de experimentos a realizar y la información que se desea obtener.
Plackett-Burman ⁶	<ul style="list-style-type: none">- Caso especial de un diseño factorial fraccionado. Es un diseño de dos niveles con un número de experimentos múltiple de 4.- Solo para estudiar efectos principales de los factores en la respuesta.
D-Óptimo ⁶	<ul style="list-style-type: none">- Permiten realizar estudios de factores con más de dos niveles sin tener el

Tabla 3.1 Diseño de experimentos para realizar estudios de *screening*.

Diseño de experimentos	Características
	<p>mismo número de niveles entre ellos.</p> <p>- Busca el mejor subconjunto de experimentos dentro de todo un conjunto, el cual maximiza el determinante de la matriz $X'X$ para un modelo de regresión.</p>

3.2. PRETRATAMIENTO DE DATOS

El pretratamiento de los datos consiste en aplicar una operación matemática al conjunto de datos antes de llevar a cabo el análisis multivariable⁷. El objetivo es aumentar o enfatizar la información de interés y reducir el ruido o información no deseada⁸. Los pretratamientos se pueden clasificar en dos grupos según la dirección en la que se aplica la operación matemática en la matriz de datos. Teniendo una matriz de datos de n muestras y m variables, los pretratamientos se pueden aplicar por columnas -variable a variable- o por filas -muestra a muestra-⁹, tal como se representa en la Figura 3.1.

Existe una gran variedad de pretratamientos, en las siguientes secciones se describen los usados en el desarrollo de esta tesis.

3.2.1. Pretratamientos por columnas

3.2.1.1. Centrado a la media

La aplicación del centrado a la media sirve para suprimir los offset presentes de manera constante en los datos¹⁰. Esto se lleva a cabo eliminando las intensidades absolutas de las variables². Para ello el sistema de coordenadas se traslada al centroide de los datos, sin que el posicionamiento entre ellos se vea afectado tal como se muestra en la Figura 3.2.

Para aplicar este pretratamiento se calcula el promedio de cada variable de la matriz de datos y esta se resta a cada una de las variables siguiendo la ecuación Ec 3.1:

$$X_{MC} = X - \bar{x} \quad \text{Ec 3.1}$$

siendo X_{MC} la matriz de datos transformados, X la matriz de datos original y \bar{x} el vector promedio de las variables.

Generalmente, este pretratamiento se aplica siempre que se realiza un modelo de calibración^{2,11}.

3.2.1.2. Escalado

El escalado se aplica siempre que las unidades de las variables tengan distintas magnitudes para evitar que las de mayor grado tengan más influencia que las de menor¹². Éste pretratamiento se aplica después del centrado a la media como se muestra en la ecuación Ec 3.2. De esta manera se obtiene para cada columna un promedio de 0 y varianza de 1.

$$X_{SC} = \frac{X_{MC}}{s} \quad \text{Ec 3.2}$$

siendo X_{SC} la matriz de datos autoescalada, X_{MC} la matriz de datos centrada en la media y s el vector de desviaciones estándar de las variables.

3.2.2. **Pretratamientos por filas**

Este tipo de pretratamientos solo se pueden aplicar en los casos que las variables sean continuas y sirven para reducir variaciones aditivas y multiplicativas³ entre muestras, reducir el ruido y enfatizar diferencias entre muestras. En esta sección se definirán los más usados en espectroscopia NIR y Raman.

3.2.2.1. Corrección del efecto multiplicativo de dispersión (MSC)

El objetivo del pretratamiento por corrección del efecto multiplicativo de dispersión (MSC del inglés *Multiplicative Scattering Correction*) consiste en reducir variaciones de la línea base de los espectros. Estas variaciones son debidas a los efectos multiplicativos y/o aditivos^{2,13} causadas por las diferencias físicas de la muestra.

Para aplicar este pretratamiento, en primer lugar se deben estimar los coeficientes de corrección multiplicativo (b_1) y aditivo (b_0) mediante una regresión lineal (ecuación Ec 3.3) usando un espectro de referencia (x_{ref}). La elección del espectro de referencia es el punto crítico de la aplicación del modelo. Se suele escoger el promedio de un conjunto de datos¹⁴.

$$x = b_0 + b_1 x_{ref} \quad \text{Ec 3.3}$$

Una vez estimados los coeficientes de corrección (tanto el multiplicativo como el aditivo), estos se usan para aplicar el pretratamiento a la matriz de datos problema (ecuación Ec 3.4).

$$x_{MSC} = \frac{x - b_0}{b_1 x_{ref}} \quad \text{Ec 3.4}$$

donde x_{MSC} es el espectro procesado y x el espectro original.

El modelo calculado para aplicar el pretratamiento MSC asume que las variaciones espectrales por efectos multiplicativos y aditivos son mayores que las debidas a variaciones químicas. En el caso de que esta premisa no se cumpliera y se aplicara este pretratamiento se podrían obtener modelos de predicción erróneos².

3.2.2.2. Variable normal estándar (SNV)

El pretratamiento de variable normal estándar (SNV del inglés *Standard Normal Variate*) se aplica para corregir los efectos aditivos y multiplicativos en la señal¹⁴. Esto efectos son debidos a diferencias físicas de la muestra que no están relacionados con la propiedad de interés que se pretende modelar. Al tratarse de un pretratamiento por filas, para cada espectro se aplica la ecuación Ec 3.5.

$$x_{SNV} = \frac{x - \bar{x}}{s} \quad \text{Ec 3.5}$$

donde x_{SNV} es el espectro corregido, \bar{x} y s el promedio y desviación estándar del espectro original respectivamente y x el espectro original.

La aplicación del SNV da resultados similares a MSC dado que los dos tienen como objetivo corregir los efectos debidos a diferencias físicas de las muestras. Aun así, a diferencia del MCS, el SNV no se calcula por ajuste de mínimos cuadrados, por lo que puede ser sensible a la presencia de ruido en el espectro¹⁵.

3.2.2.3. Derivada Savitzky-Golay

La aplicación de la derivada permite aumentar la resolución de bandas solapadas^{2,13}. Las más usadas son la primera y segunda derivada. La primera derivada reduce el desplazamiento constante

de la línea base entre muestras, mientras que la segunda derivada reduce tanto el desplazamiento constante y el proporcional de la línea base^{2,15,16} entre muestras. Una desventaja de derivar las señales es que disminuye la relación señal ruido, por lo que en general va acompañada de un proceso de suavizado¹⁵, como en el caso del método Savitzky-Golay.

Para el cálculo de la derivada de Savitzky-Golay se define un ancho de ventana móvil que abarca un número determinado de puntos del espectro. En él se ajusta un polinomio de orden determinado y se calcula la derivada en el punto central de dicha ventana. Una vez calculado el primer punto, se mueve la ventana una posición a la derecha y se repite el proceso hasta que se han procesado todos los puntos que contiene el espectro¹³.

Para poder aplicar el pretratamiento, se deben definir previamente tres parámetros: la amplitud de la ventana móvil (cuanto mayor sea, mayor el efecto de suavizado de señal), el orden del polinomio (típicamente 2 o 3) y el orden de la derivada (primera o segunda). Si el orden de derivada es 0, lo que se realiza es un suavizado de la señal¹⁵.

3.3. ANÁLISIS CUALITATIVO

El objetivo de realizar un análisis cualitativo de datos es clasificar las muestras según las variables descriptivas⁸. Este análisis puede realizarse mediante métodos no-supervisados o supervisados.

- Métodos no-supervisados: se desarrollan sin tener información previa de las muestras. Se utilizan para explorar los datos o encontrar similitudes¹⁷ entre objetos y así extraer información que no es posible obtener visualizando la matriz.
- Métodos supervisados: se desarrollan con muestras de calibración de las cuales se conoce de antemano a la clase que pertenecen. Estos tienen como objetivo poder clasificar nuevas muestras desconocidas según una regla predeterminada⁷. A la vez, los métodos supervisados se pueden clasificar en discriminantes y de modelado de clases.
 - o Métodos discriminantes: se caracterizan por dividir el espacio en tantas clases como haya en las muestras de referencia definiendo líneas fronterizas entre ellas. Una nueva muestra siempre se clasificará en una de las clases definidas.
 - o Métodos de modelado de clase: se calcula un modelo o espacio para cada clase. Una nueva muestra se identificará si se ajusta a alguno de los modelos, sino, no quedará clasificada¹⁷. Ver esquemáticamente la diferencia entre ambos métodos en la Figura 3.3.

En las siguientes secciones se van a describir los métodos cualitativos que se han usado para el desarrollo de la tesis, tanto para la exploración de datos como para su clasificación.

3.3.1. Análisis de componentes principales (PCA)

El análisis de componentes principales (PCA del inglés *Principal Component Analysis*) es una técnica lineal multivariable no supervisada que aplica una reducción de variables. La mejora computacional y tecnológica ha hecho que se disponga de grandes cantidades de datos procedentes de técnicas analíticas y/o del proceso de fabricación, por lo que la compresión de datos se hace necesaria para obtener la información relevante y descartar la redundante y el ruido.

El procedimiento matemático del PCA consiste en la descomposición de la matriz de datos compuesta por n muestras y m variables ($N \times M$) siguiendo la ecuación Ec 3.6.

$$X = T \cdot P^T + E \quad \text{Ec 3.6}$$

donde T son los *scores* ($N \times A$), P los *loadings* ($A \times M$) y E los residuales (varianza de los datos originales no explicada por el modelo).

A es el número de componentes principales (PC del inglés de *Principal Components*) que se escogen para definir el modelo². Los PC son los nuevos ejes, siendo $A \ll M$ para que se realice una reducción de datos real. Estos PC son combinaciones lineales de las variables originales y son ortogonales entre sí. El primer PC es el que define la mayor variabilidad de los datos. El segundo es

ortogonal al primero y describe la segunda mayor variabilidad, como se muestra en la Figura 3.4. Los *scores* son las proyecciones o coordenadas de las muestras en los nuevos ejes y los *loadings* son los cosenos de las variables originales con cada PC¹⁸.

El número de PC usados para calcular el modelo se deben elegir cuidadosamente siguiendo el principio de parsimonia, es decir, el modelo más simple suele ser el más robusto¹³. De esta manera se evita el sobreajuste del modelo (incorporar ruido o información irrelevante al modelo, de manera que explique muy bien los objetos usados para desarrollar el modelo pero no tenga capacidad predictiva) y también el subajuste (el modelo no explica la variabilidad de los datos).

El PCA se puede convertir en un método supervisado cuando se usa como herramienta de predicción para proyectar nuevas muestras en el espacio modelado con las muestras de referencia. A partir de los residuales del PCA se calculan dos valores estadísticos, el de Hotelling T² y el valor residual Q. El primero refleja a qué distancia del centro está una muestra dentro del espacio del modelo; mientras que el segundo, residual Q, indica la cantidad de respuesta de una muestra que se encuentra fuera del espacio PCA, que no explica el modelo³. Así, con las muestras de calibración se definen los límites de cada estadístico con un nivel de confianza determinado. Al proyectar una nueva muestra se puede detectar si tiene un comportamiento distinto al conjunto de calibración si sus residuales caen fuera de estos límites².

Las principales aplicaciones del PCA son reducir la dimensionalidad de los datos, visualizar y explorar los datos para detectar patrones, tendencias, agrupaciones, muestras anómalas y evaluar la calidad de las respuestas. También sirve como base de otros métodos de calibración multivariable, tanto cualitativos como cuantitativos.

3.3.2. Coeficiente de correlación

El coeficiente de correlación es la herramienta más simple para medir la similitud entre datos¹⁹ con un mismo número de variables. Si se comparan dos vectores, el coeficiente de correlación se entiende como el coseno del ángulo que forman ambos⁸ y se calcula según la ecuación Ec 3.7.

$$r_{x,y} = \frac{\sum_{i=1}^m x_i y_i}{\sqrt{\sum_{i=1}^m x_i^2 \cdot \sum_{i=1}^m y_i^2}} \quad \text{Ec 3.7}$$

donde r es el coeficiente de correlación entre dos vectores, x e y son dos vectores de dimensión i .

Como mayor sea r más similares serán los vectores. El coeficiente de correlación se puede usar o para un análisis de comparación de dos muestras o bien para clasificar una nueva muestra a un grupo cuando el valor calculado respecto al vector promedio de los datos de referencia (muestras de calibración) sea mayor al *threshold* (umbral) predefinido en el desarrollo del método.

3.3.3. Distancias

Las distancias son otra medida de similitud entre objetos. Estas se pueden calcular entre dos objetos o entre un objeto y un punto central. Cuanto menor sea el valor obtenido para la distancia, mayor será la similitud entre ambos. Para aplicar el cálculo de distancias, se puede trabajar en dos espacios, el de las variables originales o en el de los componentes principales (PC). Independientemente del espacio en el que se aplique las dos distancias más usadas en análisis cualitativos son la Euclídea y la de Mahalanobis.

La distancia Euclídea entre dos objetos x e y en el espacio de las variables originales se calcula según la ecuación Ec 3.8, donde m es el número de variables. La distancia entre un objeto y el punto central de una serie de objetos o una clase se obtiene según la ecuación Ec 3.9¹⁹. Si se trabaja en el espacio de PCA, se substituye el parámetro x de la ecuación Ec 3.9 por t (*scores*)².

$$d_{x,y} = \sqrt{\sum_{i=1}^m (x_i - y_i)^2} \quad \text{Ec 3.8}$$

$$d_{x,\bar{x}} = \sqrt{(x_i - \bar{x})(x_i - \bar{x})^T} \quad \text{Ec 3.9}$$

La distancia de Mahalanobis entre un objeto y el punto central de un conjunto de datos viene dado por la ecuación Ec 3.10 y tiene en cuenta que algunas variables pueden estar correlacionadas entre sí introduciendo la inversa de la matriz de varianza-covarianza (C). Una desventaja de esta distancia es que el número de muestras debe ser mayor al número de variables, dado que si no se podría dar el caso de que no se pudiera calcular la inversa de la matriz C ²⁰. Por lo que cuando se trabaja con datos con una cantidad elevada de variables, se calcula una vez aplicado el PCA, y substituyendo x por t en la ecuación Ec 3.10. Nótese que trabajando en el espacio PCA, la matriz de varianza-covarianza solo contendrá elementos diferentes a 0 en la diagonal dado que los PC son ortogonales entre sí², por lo que la distancia de Mahalanobis acaba siendo equivalente a la Euclídea²¹.

$$d_{x,\bar{x}} = \sqrt{(x_i - \bar{x})C^{-1}(x_i - \bar{x})^T} \quad \text{Ec 3.10}$$

Igual que en el caso del coeficiente de correlación, estas distancias se pueden usar para determinar la similitud entre dos objetos o bien para clasificar un objeto desconocido. Para este último caso se fijará un *threshold* con muestras conocidas. Si la distancia calculada entre el nuevo objeto y el centroide del conjunto de calibración está por debajo del valor *threshold*, significará que pertenece a la clase frente a la que se ha comparado.

3.3.4. Análisis lineal discriminante (LDA)

El análisis lineal discriminante (LDA del inglés *Linear Discriminant Analysis*) es un método de clasificación supervisado discriminante, que sirve para encontrar límites entre clases y así poder clasificar nuevas muestras.

De la misma manera que el PCA, el LDA es un método de reducción de variables²² en el que se define un nuevo espacio con un conjunto de vectores propios o funciones lineales discriminantes (LD). Estos son combinaciones lineales de las variables originales y son ortogonales entre sí. Sin embargo, a diferencia del PCA, en el caso del LDA se buscan las direcciones que maximizan la varianza entre clases y la minimizan dentro de una misma clase^{1,10,22}. En la Figura 3.5 se puede observar gráficamente la diferencia entre los PC y las LD.

El modelo de clasificación se definirá por el número de LD escogidos, evitando el sobreajuste². Sin más limitaciones, una nueva muestra se clasificará mediante la transformación de las variables originales a las nuevas coordenadas en el nuevo espacio con las funciones LD. La muestra desconocida se asignará a la clase a la cual la distancia Euclídea a su centroide sea menor¹⁰. De esta manera, y como método discriminante, un objeto siempre será asignado a una clase (a excepción que caiga justo en un límite entre clases), sin ser capaz de detectar muestras que no pertenecen a ninguna de las clases definidas en el modelo. Hay distintas formas de definir límites para prevenir esta situación. Una opción es definir un rango de varianza en cada clase, por debajo del cual una nueva muestra será clasificada como tal¹⁰.

Este método de clasificación asume que las diferencias entre clases en el espacio reducido se pueden explicar de forma lineal, por lo que si esto no se cumple, el método no funcionará

correctamente². Además, también puede presentar problemas cuando las varianzas internas de las distintas clases son muy diferentes entre sí¹⁰.

3.3.5. SIMCA

SIMCA (del inglés *Soft Independent Modelling of Class Analogies*) es un método de clasificación supervisado de modelado de clases. Es decir, para cada clase se define un modelo matemático independiente.

En este caso se calcula un modelo PCA para cada clase y se definen unos límites a un nivel de confianza determinado para los residuales de Hotelling T^2 y Q independientemente. Cuando se pretende clasificar una nueva muestra, se proyecta en todos los modelos, obteniendo tantos pares de valor T^2 y Q como clases haya en el método. Para aquellas clases en que estos valores sean menores que los definidos según el nivel de confianza, la nueva muestra será clasificada como tal. Es importante destacar que en este caso, un objeto puede ser determinado cómo perteneciente a una clase, a varias o ninguna^{2,10}

3.4. ANÁLISIS CUANTITATIVO

El análisis multivariable cuantitativo se lleva a cabo para encontrar una función que relacione una o varias respuestas cuantitativas con más de una variable predictora. Estos modelos pueden ser lineales o no-lineales, siendo los primeros los más usados debido a que la mayoría de variables dependientes y los parámetros de la regresión siguen este comportamiento. Su amplio uso también se debe a que en los casos que no se cumple este comportamiento, se puede conseguir un comportamiento lineal realizando una aproximación local o bien aplicando una transformada de los datos²².

Por otro lado, existe la diferencia entre calibraciones clásicas, donde las variables predictoras son dependientes de la respuesta, y calibración inversa, que por motivos matemáticos la respuesta actúa como variable dependiente de la matriz de variables predictoras.

Otra clasificación de los métodos cuantitativos es la que diferencia los métodos directos, donde los parámetros de regresión se determinan a partir de señales puras y los métodos indirectos donde se determinan a partir de la mezcla de señales¹⁸.

Existen diversos métodos para realizar la calibración cuantitativa como son la regresión lineal múltiple (MLR del inglés *Multiple Linear Regression*), regresión por componentes principales (PCR del inglés *Principle Component Regression*), regresión parcial por mínimos cuadrados (PLS del inglés *Partial Least Squares Regression*) o redes neuronales (ANN por sus sigla en inglés

Artificial Neural Network), siendo PLS la más usada debido a sus ventajas respecto a MLR, PCR y ANN:

- La calibración MLR es un método inverso de regresión lineal que solo se puede realizar cuando el número de objetos es mayor al número de variable³. Además, este método puede dar modelos inestables cuando hay correlaciones en las variables predictoras¹⁹ como es el caso de datos espectroscópicos.
- La calibración PCR solventa el problema de colinealidad observado en el MLR, debido a que primero se realiza un PCA en la matriz de datos predictores, y posteriormente se aplica un MLR entre la matriz de *scores* y el vector respuesta. La principal desventaja de PCR es que los primeros PC no tienen por qué estar relacionados con la variación de la respuesta².
- ANN son métodos no paramétricos de calibración no lineal³. Se caracterizan por ser métodos muy potentes, pero de una elevada complejidad de desarrollo. Debido a esta complejidad, aún no se usan de manera habitual en la industria farmacéutica.

3.4.1. Regresión parcial por mínimos cuadrados (PLS)

La regresión PLS es un método de calibración multivariable lineal indirecto, o lo que es lo mismo, los parámetros de calibración se determinan a partir de la señal que tiene la matriz, e inverso, es decir que la concentración es la variable dependiente y la señal analítica es la independiente¹⁸. También se realiza una reducción de variables, pero a diferencia del PCA, la matriz de variables (X) se descompone usando la información de la matriz respuesta (Y), ecuaciones Ec 3.11 y Ec 3.12. Las nuevas variables latentes se calculan de manera que la varianza explicada en X sea maximizada y a la vez la covarianza entre estas y la respuesta Y sea máxima².

$$X = T \cdot P^T + E \quad \text{Ec 3.11}$$

$$Y = U \cdot Q^T + F \quad \text{Ec 3.12}$$

donde T y U son los *scores* de la matriz X e Y respectivamente, P y Q los *loadings* y E y F los residuales para cada una de las matrices.

El número de factores o variables latentes que definen el modelo, como en el caso de PCA, se seleccionan siguiendo el principio de parsimonia, evitando tanto el sobreajuste como el subajuste del modelo.

Una vez realizada la descomposición se determinan los coeficientes B (parámetros de calibración) siguiendo la ecuación Ec 3.13.

$$B = W(P^T \cdot W)^{-1} \cdot Q^T \quad \text{Ec 3.13}$$

siendo W la matriz de pesos de los *loadings*.

Los parámetros de calibración B son los que permiten predecir una nueva muestra sin necesidad de transformar o descomponer las variables predictoras²³.

La calidad del modelo y de las predicciones se puede diagnosticar mediante los residuales, es decir que si un residual en Y es elevado, indican que el modelo es pobre para explicar esa respuesta. Por otro lado, los residuales en X son la parte no utilizada para modelar Y ²⁴.

3.5. DESARROLLO DE UN MÉTODO NIR Y RAMAN EN EL ENTORNO FARMACÉUTICO

El desarrollo de un modelo de calibración es el proceso para relacionar una o varias propiedades de las muestras con las variables predictoras, siendo estas los espectros cuando se trabaja con NIR o Raman. El proceso de calibración consta de distintas etapas descritas en la Figura 3.6.

Cuando el objetivo es desarrollar un método en la industria farmacéutica es indispensable que se haga dentro del marco regulador correspondiente. Para ello, se deberán seguir las guías publicadas por las agencias regulatorias EMA y FDA sobre el desarrollo de un método NIR^{25,26} (extensibles a cualquier técnica espectroscópica que requiera el empleo de quimiometría para poder ser aplicadas²⁷), así como las indicaciones sobre validaciones de método analítico ICH Q2²⁸. Por último, también sirven de soporte los capítulos publicados por la farmacopea europea y estadounidense sobre quimiometría^{19,29}, NIR³⁰ y Raman³¹.

A continuación se describen las distintas etapas para el desarrollo de un método de calibración NIR o Raman según la Figura 3.6.

3.5.1. Definición del objetivo

Inicialmente se debe definir de forma clara cuál será el objetivo del método analítico a desarrollar, así como lo esperado de dicho método. Para ello, se debe disponer de toda la información relacionada con el método, como:

- El objetivo del método.
- El tipo de tecnología que se usará para el análisis.
- Si se pretende sustituir un método utilizado o ampliar el conocimiento del proceso/producto. En el caso de que se quiera sustituir un método, si este está en el expediente de registro autorizado del medicamento.
- Si el análisis se pretende llevar a cabo en el laboratorio de control de calidad o en la planta de producción. En el último caso, se realizaría con un equipo *on-*, *in-* o *at-line*.
- Realizar un análisis de riesgos para identificar todas las variables que pueden afectar al análisis.

Una vez se disponga de toda esta información se debe establecer un protocolo para diseñar la estrategia de desarrollo y validación del método analítico.

3.5.2. Preparación del conjunto de calibración y validación

Esta es una de las etapas claves para obtener un modelo de calibración con mayor exactitud y robustez. Para conseguirlo, se debe introducir todas las fuentes de variabilidad esperadas en las muestras que se pretenden predecir en un futuro, así como, en el caso de métodos cuantitativos, expandir el rango de la propiedad de interés que se pretende calcular. Por otro lado, es necesario ser cuidadoso con la variabilidad que se introduce dado que pueden aparecer no linealidades por introducir demasiada en el sistema³².

En el caso de productos farmacéuticos, dependiendo del tipo de calibración que se vaya a realizar, la variabilidad se introducirá como sigue:

- Variabilidad física (métodos cualitativos y cuantitativos): introduciendo muestras elaboradas con componentes de distintos proveedores, tamaños de partícula, equipos de producción...
- Variabilidad química (métodos cuantitativos): esta se hará expandiendo el rango de la propiedad de interés. Cuando esta propiedad es la concentración de algún componente, no se puede realizar mediante el uso de distintos lotes de producción, puesto que en la

industria solo se pueden encontrar variaciones de concentración de API del $\pm 5\%$ respecto al valor nominal. Este rango es demasiado estrecho, si lo comparamos con las directrices de la ICH Q2, donde indica que, para una riqueza el rango debe ser $\pm 20\%$ mientras que para una uniformidad de dosis debe ser de $\pm 30\%$ ²⁸. Consecuentemente, se debe utilizar alguna de las siguientes estrategias para conseguir expandir la propiedad de interés³³:

- Planta piloto: siguiendo el mismo proceso de fabricación en pequeña escala que en la planta de producción permitiendo introducir tanto la variabilidad química como la física. Sin embargo, requiere de un coste elevado y muchas veces se produce más cantidad de muestra que la necesaria para crear el conjunto de calibración.
- Muestras de laboratorio: se basa en la mezcla de los componentes de la formulación en polvo. Esto permite expandir la variabilidad química, pero físicamente son distintas a las muestras de producción, por lo que en el conjunto de calibración también se deben incluir muestras de producción.
- Muestras dopadas: se trata de sobre-dosificar y sub-dosificar muestras de producción usando API o placebo. De esta manera se introduce la variabilidad química a la vez que se introducen las características de las muestras de producción.

Para asegurar que la variabilidad espectral se correlacione con la propiedad de interés y no con otras³⁴ -como por ejemplo con la variación de excipientes en vez de con el API- es necesario realizar un DoE y así intentar reducir correlaciones entre el API y los excipientes³⁵.

Las muestras preparadas formarán parte de dos conjuntos, el de calibración y el de validación, y deben ser independientes entre sí. El conjunto de calibración se usará para desarrollar el modelo y optimizarlo, mientras que el conjunto de validación servirá para evaluar la capacidad predictiva del modelo.

3.5.3. Adquisición del espectro

La adquisición de los espectros se debe realizar con un equipo que haya sido cualificado según el Anexo 15 de las GMP. Además, en equipos como el NIR o el Raman, se realizan verificaciones periódicas, usualmente diarias, siguiendo el capítulo genérico de la farmacopea europea para ambas técnicas^{30,31}, para asegurar el correcto funcionamiento de los mismos.

Siguiendo lo explicado en el capítulo 2, el modo de adquisición -reflectancia o transmisión- y los parámetros de adquisición de los espectros, deben ser evaluados y escogidos según los objetivos del análisis y el tipo de muestra a analizar.

Los espectros obtenidos de todas las muestras se visualizan para observar el comportamiento espectral, así como para localizar espectros anómalos. Los espectros anómalos que no contengan información de la muestra se deben eliminar antes de proceder con los siguientes pasos, realizando una justificación en el informe final de la validación del método.

3.5.4. Determinación de la propiedad por el método de referencia

Para ambas técnicas la señal analítica no varía únicamente debido a la propiedad de interés, hay otros factores que influyen en ella como son la matriz, las propiedades físicas de la muestra, las condiciones ambientales, el instrumento... Por lo que hablamos de técnicas analíticas secundarias, esto es, deben ser calibradas frente a unos valores obtenidos por un método de referencia³⁶. Estos deben ser los que están aprobados en el expediente. Se debe analizar la misma muestra por NIR o Raman y por el método de referencia.

En el casos de análisis cuantitativos es necesario identificar el error estándar de laboratorio (SEL del inglés *Standar Error of Laboratory*), siendo la precisión intermedia o la reproducibilidad del método de referencia²⁵.

3.5.5. Pretratamiento de los datos y selección rango espectral

El pretratamiento se aplica con el fin de reducir interferencias provocadas por las características de la muestra o por el instrumento en el espectro. Estas interferencias no están directamente relacionadas con las propiedades de interés. En el caso del NIR se deben principalmente a la dispersión de la luz, el camino óptico, las propiedades físicas de la muestra...mientras que en el caso del Raman se deben a la fluorescencia, fluctuación del láser.... En los espectros, estas interferencias se suelen presentar como factores aditivos, factores multiplicativos, desplazamiento polinomial de la línea base o ruido³⁷. Para este fin se usarán los métodos mencionados en el apartado 3.2. *Pretratamiento de datos*.

Al tratarse de métodos empíricos no hay una forma guiada de proceder. El mejor pretratamiento o combinación de ellos será aquel con el que se obtenga el modelo más robusto sin eliminar la información de interés¹⁵ y sin introducir nuevas variaciones en los datos¹⁶. La elección se suele realizar mediante el método de prueba y error, comparando el error cuadrático medio de predicción (RMSEP del inglés *Root Mean Squared Error Prediction*) o el error cuadrático medio de validación cruzada (RMSECV del inglés *Root Mean Squared Error Cross-Validation*).

Otro parámetro a optimizar de manera iterativa es el rango espectral a usar, pudiendo ser completo o seleccionando uno específico. La segunda opción puede ayudar a mejorar las predicciones y la robustez del modelo³⁸, pero también puede disminuir la capacidad del modelo para detectar *outliers*³⁹.

3.5.6. Exploración de los datos y detección de *outliers*

Para visualizar los datos y comprobar el comportamiento de los espectros es útil realizar un análisis exploratorio de los datos en el espacio multivariable mediante la herramienta PCA.

El gráfico de *scores*, representación de las coordenadas de cada muestra en los nuevos ejes (PC), es útil para estudiar el comportamiento de las muestras (detectar tendencias, grupos entre muestras...).

El PCA también puede ser útil para detectar o identificar posibles *outliers* espectrales. Un *outlier* es una muestra que se comporta de manera distinta a la población a la que pertenece⁷. Pueden existir distintos tipos como son los relacionados con el espectro, con el valor de referencia o con la relación entre el espectro y el valor de referencia. Un *outlier* puede estar provocado por una mala medición, pero también puede pertenecer a una variabilidad inherente a la muestra, por lo que no tiene por qué ser incorrecto².

Los *outliers* espectrales se pueden detectar mediante el gráfico de los residuales Q frente al valor de Hotelling T^2 ¹³ obtenidos con el cálculo del PCA. Una vez detectado un *outlier* se debe elegir y justificar si se elimina o se mantiene en el conjunto de muestras.

3.5.7. Selección de muestras

Si se dispone de un número suficiente de muestras de calibración, estas se van a dividir en el conjunto de calibración y el de optimización del modelo³⁴. Para realizar esta selección existen varios métodos, ejemplo de ellos son:

- Manual: a partir del gráfico de *scores* del PCA realizado para explorar de los datos se seleccionan manualmente las muestras de manera que el conjunto de calibración englobe toda la variabilidad posible.
- Kennard-Stone: algoritmo que normalmente se aplica en el espacio PCA y que permite dividir las muestras en conjunto de calibración y optimización. Este empieza con la selección de las dos muestras más distintas entre ellas, la siguiente muestra escogida es la que está más lejos de ellas. Este proceso se repite hasta que el número de muestras es el mismo que el indicado inicialmente¹³.

Por contra, si no se dispone de suficientes muestras, la optimización del modelo se realizará mediante la validación cruzada. Este método consiste en usar las mismas muestras empleadas para la calibración del modelo para su optimización. Su funcionamiento se basa en calcular el modelo con una parte de las muestras dejando cierta cantidad fuera. Con el modelo obtenido, se predicen las muestras que se han dejado fuera y se calculan los residuales. Este proceso se repite con otro subconjunto, hasta que al final, todas las muestras se han dejado fuera al menos una vez. La forma

de dejar muestras fuera del cálculo durante la validación cruzado se hace de manera sistemática y depende del método que se seleccione¹³. La elección de dicho método se debe definir.

3.5.8. Modelo de calibración y optimización

Tanto en modelos cualitativos como cuantitativos la elección del algoritmo a usar para realizar la calibración del modelo se hará según el objetivo del estudio.

3.5.8.1. Método cualitativo

El primer paso consiste en decidir el espacio que se va a usar, esto es, el espectral o una reducción de variables. Según esta elección, se escogerá el algoritmo matemático para desarrollar el modelo. En el caso de que el método sea de reducción de variables, la elección del número de factores se realizará con el conjunto de muestras de optimización. Usando el gráfico de varianza explicada frente al número de los nuevos ejes (PC, LD...), se escogerá aquel a partir de cual la varianza explicada no mejore significativamente introduciendo un nuevo factor.

Mediante la validación interna o el conjunto de optimización se determinará una regla de clasificación y se evaluará el funcionamiento del modelo asegurando su selectividad.

3.5.8.2. Método cuantitativo

El PLS es el algoritmo matemático más usado para calibrar modelos cuantitativos debido a sus ventajas. La elección del número de factores para definir el modelo se realizará con el conjunto de optimización. Usando el gráfico de los valores RMSECV o de RMSEP frente al número de factores se escogerá aquel a partir de cual el residual no mejore significativamente introduciendo un nuevo factor⁴⁰. No se usa el de calibración puesto que siempre mejora.

$$RMSECV/RMSEP = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}} \quad \text{Ec 3.14}$$

donde y_i es el valor de referencia, \hat{y}_i el valor predicho por el modelo y N el número de muestras del conjunto de validación cruzada (RMSECV) y de predicción (RMSEP). Los valores RMSE tienen las mismas unidades que la respuesta que se predice. El cálculo del error estándar relativo (RSE del inglés *Root Standard Error*) de la ecuación Ec 3.15 permite obtener valores del error adimensionales.

$$RSECV/RSEP = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i)^2}} \cdot 100 \quad \text{Ec 3.15}$$

donde y_i es el valor de referencia, \hat{y}_i el valor predicho por el modelo y N el número de muestras del conjunto de validación cruzada (RSECV) y de predicción (RSEP).

3.5.9. Validación externa

La validación del método se lleva a cabo según las indicaciones de la guía de la EMA sobre NIR²⁵ y de la ICH Q2²⁸ con el conjunto de validación externa, que es un grupo de muestras independientes de las usadas durante la calibración y optimización de los modelos. En la Tabla 3.2 se indican los parámetros que se deben determinar para métodos cualitativos y cuantitativos.

Tabla 3.2 Parámetros a determinar durante el proceso de validación de métodos cualitativos y cuantitativos.

Parámetro	Método cualitativo	Método cuantitativo
Especificidad	La especificidad se define como “la habilidad de evaluar de manera inequívoca el analito en presencia de componentes que pueden estar presentes” ²⁸ . Es necesario demostrar que no hay falsos positivos ni negativos y que es capaz de clasificar correctamente las muestras que pertenecen a una clase incluida en el modelo.	Se suele realizar una identificación antes de proceder a cuantificar para asegurar que la nueva muestra pertenece a la clase y a la población del conjunto de calibración usado para realizar el modelo. Además, también se pueden usar los gráficos de <i>loadings</i> para demostrar que los factores que se han incluido en el modelo están relacionados con la propiedad de interés (las bandas principales del compuesto de interés están presentes en los <i>loadings</i> de cada factor).
Linealidad y rango	No aplica	La linealidad se define como la “habilidad (dentro de un rango concreto) de obtener resultados que son directamente proporcionales a la propiedad de interés en la muestra” ²⁸ . Para evaluarla, será necesario realizar una regresión lineal entre los valores predichos por el modelo y los valores de referencia obtenidos en el conjunto de validación. El intervalo de confianza la pendiente y ordenada en el origen debe contener el 1 y el 0 respectivamente con un nivel de confianza del 95%. También se debe comprobar la homocedasticidad de los residuales frente al valor de la propiedad de interés. El rango queda definido por el valor máximo y mínimo de la propiedad con los que se ha evaluado la linealidad.
Exactitud	No aplica	La exactitud del modelo es “el grado de coincidencia entre el valor que se considera verdadero y el valor encontrado por el procedimiento analítico” ²⁸ . Para ello, se realizará un test t de muestras emparejadas entre el valor predicho y el valor de referencia de las muestras usadas para la linealidad, de manera que se aceptará el modelo solo si el valor t calculado es menor al valor t crítico de dos colas con un 95% de nivel de confianza.
Precisión	No aplica	La precisión se define como “el grado de coincidencia entre una serie de medidas obtenidas de múltiples muestreos de una misma muestra homogénea bajo condiciones determinadas” ²⁸ . Se

Tabla 3.2 Parámetros a determinar durante el proceso de validación de métodos cualitativos y cuantitativos.

Parámetro	Método cualitativo	Método cuantitativo
		determinarán dos niveles, repetitividad y precisión intermedia. - Repetitividad Precisión de un método bajo las mismas condiciones de medida para un intervalo corto de tiempo ²⁸ . Para evaluarlo se realizan o bien 9 determinaciones cubriendo el rango del modelo o bien 6 al valor nominal. - Precisión intermedia En este caso se analiza una misma muestra durante tres días por dos analistas diferentes. Se realiza un análisis de varianza (ANOVA) de dos factores para verificar que no hay efecto de estos dos factores sobre la predicción.
Límite de detección / cuantificación	No aplica	El límite de detección y cuantificación es “la mínima cantidad de analito en una muestra que se puede detectar y cuantificar respectivamente” ²⁸ . Estos solo se determinarán en los casos en que la propiedad de interés se considere una impureza.
Robustez	Se trata de demostrar la capacidad del método de predecir correctamente la propiedad de interés en variaciones de los parámetros de medida. Para ello se realiza un test t de muestras emparejadas de muestras adquiridas durante un periodo más o menos amplio de tiempo que contenga muestras en las que hay incorporadas todas las variabilidades de proceso, instrumento, condiciones ambientales... que permita asegurar la robustez.	

3.5.10. Uso y mantenimiento del modelo

Una vez validado el modelo, y si es el caso, aprobado por el órgano regulador, el método se puede usar en rutina.

Los modelos de calibración se deben mantener durante el ciclo de vida debido a su sensibilidad en el cambio de parámetros de materia prima, de proceso, de instrumento... Por esta razón es necesario comprobar periódicamente su validez. Para llevarlo a cabo se suelen analizar muestras con una frecuencia determinada por ambos métodos, NIR/Raman y el de referencia. Los residuales obtenidos se usarán para detectar tendencias o fuera de control mediante gráficos de control. Si se detecta que ha habido un cambio o una deficiencia en la predicción del modelo, este se deberá recalcular con muestras que contengan la nueva variabilidad, repitiendo todo el proceso de desarrollo de la Figura 3.6.

En el caso de modelos calculados con el algoritmo PLS, también es útil comprobar que los residuales T^2 de Hotelling y/o los residuales Q de las nuevas muestras se encuentren dentro de los límites de confianza definidos por el modelo. Se puede dar la circunstancia que los residuales de la

diferencia entre el valor predicho y el de referencia no presenten anomalías, pero se puedan detectar *outliers* o tendencias con los residuales T^2 y Q ⁴¹.

3.6. BIBLIOGRAFÍA

1. Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y. & Kaufman, L. *Chemometrics: A Textbook*. (Elsevier, 1988).
2. Miller, C. E. Chemometrics in Process Analytical Chemistry. in *Process Analytical Technology* 354–438 (John Wiley & Sons Ltd., 2010).
3. Roussel, S., Preys, S., Chauchard, F. & Lallemand, J. Multivariate Data Analysis (Chemometrics). in *Process Analytical Technology for the Food Industry* (eds. O'Donnell, C. P., Fagan, C. & Cullen, P. J.) 7–59 (Springer, 2014).
4. Lundstedt, T. *et al.* Experimental Design and Optimization. *Chemom. Intell. Lab. Syst.* **42**, 3.40 (1998).
5. Leardi, R. Experimental design in chemistry: A tutorial. *Anal. Chim. Acta* **652**, 161–172 (2009).
6. Johansson, E., Kettaneh-Wold, N., Wikstrom, C., Wold, S. & Ericksson, L. *Design of Experiments - Principles and Applications*. (2000).
7. Frank, I. E. & Todeschini, R. *The Data Analysis Handbook*. (Elsevier Science B.V., 1994).
8. Roggo, Y. *et al.* A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* **44**, 683–700 (2007).
9. Kramer, R. *Chemometric Techniques for Quantitative Analysis*. (CRC Press, 1998).
10. Otto, M. *Chemometrics: Statistics and Computer Application in Analytical Chemistry*. (Wiley-VCH, 2016).
11. O'Donnell, C. P., Fagan, C. & Cullen, P. J. *Process Analytical technology for the Food Industry*. (Springer, 2014).
12. Gemperline, P. *Practical guide to chemometrics*. (CRC Taylor & Francis, 2006).
13. Swarbrick, B. & Westad, F. An overview of chemometrics for the engineering and measurement sciences. in *Handbook of measurement in Science and Engineering* (ed. Kutz, M.) 2309–2407 (John Wiley & Sons Inc., 2016).
14. Rinnan, Å. *et al.* Data Pre-Processing. in *Infrared Spectroscopy for Food Quality Analysis and Control* 29–50 (Elsevier Inc., 2009).
15. Rinnan, Å., Berg, F. van den & Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends Anal. Chem.* **28**, 1201–1222 (2009).
16. Engel, J. *et al.* Breaking with trends in pre-processing? *TrAC - Trends Anal. Chem.* **50**, 96–106 (2013).
17. Ballabio, D. & Todeschini, R. Multivariate Classification for Qualitative Analysis. in *Infrared Spectroscopy for Food Quality Analysis and Control* 83–104 (Elsevier Inc., 2009).
18. Coello, J. & Maspocho, S. Calibración multivariable. in *Temas avanzados que quimiometría* 239–286 (Edicions UIB, 2007).
19. European Pharmacopoeia. Ph. Eur. 8.7 Chapter 5.21. Chemometric methods applied to analytical data. (2016).

20. Brereton, R. G. *Chemometrics Data Analysis for the Laboratory and Chemical Plant*. (Wiley, 2003).
21. Deza, M. M. & Deza, E. *Encyclopedia of distances. Encyclopedia of Distances* (Springer, 2009).
22. Vandeginste, B. G. M. *et al. Handbook of Chemometrics and Qualimetrics: Part B. Data Handling in Science and Technology* **20B**, (Elsevier, 1998).
23. Blanco Romía, M. & Alcalá Bernàrdez, M. Multivariate Calibration for Quantitative Analysis. in *Infrared Spectroscopy for Food Quality Analysis and Control* 51–82 (Elsevier Inc., 2009).
24. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).
25. EMA. Guideline on the use of Near Infrared Spectroscopy (NIRS) by the pharmaceutical industry and the data requirements for new submissions and variations. *Eur. Med. Agency* **44**, 1–28 (2014).
26. FDA. Development and Submission of Near Infrared Analytical Procedures Guidance for Industry. 24 (2015).
27. Andrews, D. *et al.* Analytical Method Development Using Transmission Raman Spectroscopy for Pharmaceutical Assays and Compliance with Regulatory Guidelines—Part I: Transmission Raman Spectroscopy and Method Development. *J. Pharm. Innov.* **13**, 121–132 (2018).
28. ICH (International Conference for Harmonisation). ICH Q2 (R1) Validation of Analytical Procedures : Text and Methodology. *International Conference on Harmonization* 17 (2005).
29. United States Pharmacopeia. USP 40–NF 35 <1039> Chemometrics. (2016).
30. European Pharmacopoeia. Ph. Eur. 9.0 Chapter 2.2.40. Near-infrared spectroscopy. (2017).
31. European Pharmacopoeia. Ph. Eur. 9.0 Chapter 2.2.48. Raman spectroscopy. (2017).
32. Næs, T. & Isaksson, T. Selection of Samples for Calibration in Near-Infrared Spectroscopy. Part I: General Principles Illustrated by Examples. *Appl. Spectrosc.* **43**, 328–335 (1989).
33. Blanco, M., Romero, M. A. & Alcalá, M. Strategies for constructing the calibration set for a near infrared spectroscopic quantitation method. *Talanta* **64**, 597–602 (2004).
34. Burns, D. A. & Ciurczak, E. W. *Handbook of Near-Infrared Analysis*. (CRC Press Taylor and Francis, 2008).
35. Romañach, R. J., Román-Ospino, A. D. & Alcalá, M. A Procedure for Developing Quantitative Near Infrared (NIR) Methods for Pharmaceutical Products. in *Process Simulation and Data Modeling in Solid Oral Drug Development and Manufacture* 133–158 (2016).
36. Osborne, B. G. Near-Infrared Spectroscopy in Food Analysis. *Encycl. Anal. Chem.* 1–14 (2000).
37. Bi, Y. *et al.* A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. *Anal. Chim. Acta* **909**, 30–40 (2016).
38. Blanco, M. & Alcalá, M. NIR Spectroscopy in Pharmaceutical Analysis: Off-Line and At-Line PAT Applications. in *Process Analytical Technology* 463–491 (2010).
39. Kjeldahl, K. & Bro, R. Some common misunderstandings in chemometrics. *J. Chemom.* **24**, 558–564 (2010).
40. De Bleye, C. *et al.* Critical review of near-infrared spectroscopic methods validations in pharmaceutical applications. *J. Pharm. Biomed. Anal.* **69**, 125–132 (2012).
41. Small, G. W. Chemometrics and near-infrared spectroscopy: Avoiding the pitfalls. *TrAC - Trends Anal. Chem.* **25**, 1057–1066 (2006).

Parte II

MÉTODOS DE IDENTIFICACIÓN

4. IDENTIFICACIÓN DE MATERIAL DE PARTIDA

4.1. INTRODUCCIÓN

Identificar el contenido de todos los envases que constituyen un lote de material de partida antes de usarse en producción es de obligado cumplimiento, excepto en algún caso especial¹. Se consideran material de partida tanto los principios activos (API por sus siglas en inglés *Active Pharmaceutical Ingredient*) como los excipientes. Una vez se han identificado correctamente, la calidad del lote se puede evaluar con una muestra representativa, el muestreo de la cual se realizará mediante criterios estadísticos². Este procedimiento se realiza para asegurar que todo el material de partida usado en la fabricación de un medicamento corresponde a lo que certifica el proveedor, y así, evitar problemas de salud pública como, por ejemplo, los que ocurrieron en Panamá el año 2006. Después de una serie de errores en la cadena de suministro, se fabricó jarabe para el resfriado usando como excipiente dietilenglicol (DEG) en vez de glicerol. Esto provocó la muerte de decenas de pacientes³⁻⁵. El glicerol o glicerina es un excipiente usado como edulcorante, agente antimicrobiano o como agente de carga para aumentar la viscosidad en soluciones líquidas orales⁶. El DEG es un sustituto barato del glicerol, que se usa en distintas industrias, pero es tóxico para el ser humano. Esta tragedia, se podría haber evitado si cada envase se hubiera identificado antes de entrar en la planta de producción.

La espectroscopia NIR y Raman son técnicas ampliamente usadas desde hace años para este fin⁷, debido a las ventajas que presentan comparado con otros métodos de identificación que se recomiendan en las farmacopeas, como son métodos clásicos (p.ej. reacciones específicas o test de precipitación) o técnicas instrumentales (p.ej. IR, HPLC o GC)⁸. Permiten el análisis sin manipular el producto, siempre que el envase primario sea transparente al fenómeno que se está midiendo. El análisis se puede realizar de manera *at-line*, ya que los instrumentos que existen en el mercado disponen de sondas o son portátiles. Esto permite realizar el análisis en la cabina de muestreo o incluso en el almacén, si es posible identificar a través del envase. Otra ventaja es la rapidez con la que se obtiene la respuesta, que permite determinar la validez de un lote de manera inmediata. Por último, al ser métodos no destructivos, la identificación se puede llevar a cabo directamente sobre el producto. En este último caso, hay una excepción con la técnica Raman. Se debe tener especial cuidado con la potencia del láser, sobre todo cuando las muestras son coloreadas, para evitar que se produzca una degradación térmica⁹. Las dos técnicas se consideran complementarias, debido a los distintos mecanismos que provocan a la señal^{7,10}.

Como se ha comentado en la introducción, las propiedades químicas y físicas de los materiales de partida suelen ser parámetros críticos de calidad, porque pueden afectar a la fabricación del

producto, así como a la calidad final del mismo¹¹. Estas técnicas, aparte de ser útiles para la identificación, permiten diferenciar materiales según sus propiedades. El NIR es capaz de diferenciar materiales químicamente idénticos pero con distinto tamaño de partícula, así como diferenciar formas polimorfas de los sólidos o viscosidades de geles y líquidos. También es una técnica sensible a la humedad, por lo que se puede determinar el contenido en agua del material. Por otro lado, el Raman puede ser útil para diferenciar formas polimórficas¹¹ y detectar la presencia de impurezas que provocan efecto de fluorescencia. En consecuencia, son técnicas que permiten realizar la cualificación de CMAs (tanto propiedades químicas como físicas) de los API y excipientes y que están directamente relacionados con las CQAs del producto terminado. Mediante la identificación se confirma la identidad química del material, mientras que con la cualificación se confirma la identidad física del mismo¹².

Es indispensable el uso de la quimiometría para aplicar ambas técnicas, NIR y Raman. Para identificar y cualificar muestras se emplean métodos de reconocimiento de patrones (PRM por sus siglas en inglés *Pattern Recognition Methods*). Se suelen desarrollar bibliotecas espectrales que contienen todos los materiales de partida y así se evita tener un método distinto para cada uno de ellos^{8,12}. En el caso del NIR, se empieza elaborando una biblioteca general que permita identificar el mayor número de sustancias posibles. El criterio de similitud que habitualmente se emplea para esta primera criba es el coeficiente de correlación o el cálculo de distancias espectrales. Se debe establecer un criterio de aceptación por el cual una muestra se considerará correctamente identificada^{5,8}. Para los productos que presentan ambigüedades entre sí, esto es, no hay selectividad entre ellos, se construyen sub-bibliotecas con un método cualitativo de mayor selectividad que permita diferenciarlas. En la Figura 4.1 se muestra un esquema de lo que se conoce como una biblioteca de identificación en cascada. En el caso del Raman, al tratarse de espectros bandas bien resueltas¹³, las bibliotecas espectrales suelen ser más sencillas, incluyendo solo una de principal.

En Laboratorios Menarini, desde el año 2001, la identificación de material partida se lleva a cabo mediante NIR. Esta identificación se realiza en una sala de muestreo, de manera *at-line*. Hasta el momento de la realización de este trabajo, la versión de la biblioteca espectral NIR contenía 184 sustancias. El objetivo de la primera parte de este trabajo ha sido actualizar la biblioteca NIR con nuevas materias primas no incluidas en la biblioteca y mejorar la robustez de las ya existentes, incluyendo nueva variabilidad no considerada en la primera versión. Por necesidades internas de la compañía, durante la estancia del doctorado se adquirió un instrumento Raman portátil para la identificación de material de partida. El segundo objetivo ha sido definir el protocolo de desarrollo y validación de una biblioteca de identificación con el equipo Raman.

4.2. MÉTODO EXPERIMENTAL

4.2.1. Muestras

4.2.1.1. NIR

Los materiales de partida incluidos en la biblioteca NIR fueron 224, englobando 65 APIs y 159 excipientes. Todas las muestras usadas para la calibración y la validación del modelo fueron identificadas por el método de referencia pertinente. El número total de lotes disponibles para cada una de ellas fue muy variable, ya que esto depende de tres factores. La cantidad usada para la preparación de los correspondientes preparados farmacéuticos, el tamaño de lote del material de partida y la cantidad del mismo por envase y del volumen de producción de cada producto terminado.

4.2.1.2. Raman

Para establecer la metodología del desarrollo de la biblioteca de identificación por Raman, se usaron muestras que permitieran verificar la técnica para así establecer la criticidad de las propiedades relacionadas con los materiales de partida. Para ello se usaron distintos APIs, excipientes o productos intermedios que presentaban una variedad de propiedades que podrían verse afectadas por la radiación incidente. Entre los que se encontraban un material que tenía dos fabricantes homologados, productos con distinto tamaño de partícula, distinta pureza o diferentes grados de polimorfismo.

4.2.2. Adquisición de espectros

4.2.2.1. NIR

Los espectros NIR se registraron con el instrumento FT-NIR MPA (Bruker Optics, Alemania) equipado con un detector InGaAs, una sonda de reflectancia difusa unida al equipo mediante una fibra óptica de 2m y un compartimento para las medidas de transmisión con un camino óptico de 5mm. El equipo se controlaba con el software OPUS 7.0. Cada espectro se adquirió en el rango

espectral 12800-4000 cm^{-1} , con una resolución de 16cm^{-1} y promediando 32 *scans*. Este equipo estaba cualificado según el fabricante y se realizaba una verificación periódica del mismo.

Las muestras sólidas se analizaron con la sonda de reflectancia difusa. En el caso de materias primas entregadas con una bolsa de polietileno transparente como envase primario, los espectros fueron adquiridos a través del plástico. El espectro de referencia se realizó con un trozo de bolsa colocada sobre la placa de cerámica del equipo. Por otro lado, si el *packaging* primario de la materia prima no permitía adquirir los espectros a través de él (bolsas de cartón, aluminio, plástico opaco...), el espectro se registró directamente sobre el producto. En consecuencia, el espectro de referencia se obtuvo directamente con la cerámica del equipo.

En el caso de las muestras semisólidas (ceras, parafinas...), se calentaron para que estuvieran en un estado líquido. Estas, junto a las muestras líquidas, se registraron en modo de transmisión. Se abrió cada contenedor para muestrearlo y las cubetas se colocaron en el compartimento de muestras del instrumento. La referencia se registró con una cubeta vacía entre el haz de luz y el detector del equipo.

4.2.2.2. Raman

Se usó en instrumento de retrodispersión portátil TruScan RM (Thermo Scientific, USA) equipado con un láser de excitación de 785nm con potencia de 250mW y un detector de Silicio CCD de 2048 píxeles. El espectro se registró en el rango de desplazamiento Raman 250-2875 cm^{-1} con una resolución variante en todo el rango de 8.0-10.5 cm^{-1} . Para el análisis de muestras sólidas se utilizó un cono auxiliar, tanto si la adquisición se hacía directamente sobre el producto o a través de una bolsa de polietileno. Para la adquisición de espectros de muestras líquidas se usaron cubetas. El Raman se controla directamente desde el instrumento, a través del *display* de la pantalla, con el software instalado TruScan Software v 2.6.2.

4.2.3. **Desarrollo de la biblioteca NIR**

4.2.3.1. Selección del conjunto de calibración

Los espectros que forman parte del conjunto de calibración, que servirán para desarrollar el método, deben contener toda la variabilidad que se espera del material de partida objetivo. Esta variabilidad proviene, en realidad, de tres fuentes. De la variabilidad presente entre adquisiciones de espectros, la variabilidad intra-lote, y la inter-lote (diferentes parámetros como son fabricantes, propiedades químicas y físicas, contenido en agua...)⁵. La señal NIR es sensible a esta variabilidad, por lo que es necesario incluirla en el conjunto de calibración. No hay una regla general que indique el número de lotes y espectros a usar para este propósito. Algunos autores recomiendan que una buena aproximación es que para realizar una identificación se usen 3 lotes y para la

cualificación de un producto sean más de 20 ¹². Como estrategia inicial se siguieron las indicaciones de la Tabla 4.1, según el número disponible de lotes para cada material.

Tabla 4.1 Criterio de elección de espectros para el conjunto de calibración y validación de la biblioteca NIR según el número de lotes disponible.

Número de lotes	Calibración	Validación
>3	3 lotes Se eligen los espectros que presentan mayor variabilidad	Resto de lotes Se usan todos los espectros
3	2 lotes Se eligen los espectros que presentan mayor variabilidad	1 lote Se usan todos los espectros
2	1 lote Se eligen los espectros que presentan mayor variabilidad	1 lote Se usan todos los espectros
1	1 lote Se eligen los espectros que presentan mayor variabilidad	Se usan los espectros no incluidos en la calibración

Si para un producto se disponía de más de tres lotes, la elección de los que formaban parte de la calibración se realizó de acuerdo a las premisas de introducir toda la variabilidad que cabe esperar. Ahora bien, en aquellos casos donde había tres o menos lotes, no era posible tener la certeza de que el método incluyera toda la variabilidad que puede presentar el producto. A excepción de aquellos productos con un solo lote, en el resto de materiales de partida los lotes de calibración y validación externa eran distintos.

Una vez escogidos los lotes de calibración, no se usaron todos los espectros que pertenecían a un mismo lote para realizar el cálculo, debido a que cada uno podía contener más de 200. Inicialmente se seleccionaron visualmente tres espectros de cada lote para formar parte del conjunto de calibración. Previamente se aplicó el pretratamiento SNV a los espectros de cada lote para reducir las diferencias físicas, y se eligieron los tres que representaban la variabilidad del lote.

Si durante el desarrollo de la biblioteca se detectaba que no se había incluido toda la variabilidad de un producto, se incluían más lotes o más espectros de un lote en el conjunto de calibración. Esta etapa era un proceso iterativo

4.2.3.2. Calibración y optimización de la biblioteca

Para desarrollar la biblioteca se utilizó el aplicativo IDENT que incorpora el software Opus 7.0. El conjunto de calibración se empleó para calcular y optimizar la biblioteca NIR. Los parámetros que se deben escoger y/u optimizar son el algoritmo usado para medir la similitud, el pretratamiento de los datos, el rango espectral incluido y, finalmente, el valor de aceptación. Los dos algoritmos de cálculo que se usaron fueron la distancia Euclidiana en el espacio espectral, y la distancia

Euclidiana en el espacio de los componentes principales (factores según el software del equipo NIR). El primero calcula la distancia entre un espectro y el espectro de referencia en el espacio de números de onda, el segundo calcula la distancia de los *scores* en el espacio reducido de componentes principales (PC). Los pretratamientos disponibles son la SNV, la primera y segunda derivada con una ventana móvil de 5 a 25 puntos, y la combinación de derivada y SNV. Estos deben servir para reducir las diferencias entre espectros de la misma sustancia y a la vez aumentar las diferencias entre espectros de distintos productos. El rango espectral se ajusta de manera que contenga tanta información como sea posible y a la vez se excluyan zonas con ruido o sin información útil. Finalmente, la elección del valor de aceptación o umbral se tiene que definir de manera cuidadosa, ya que es el valor mínimo requerido para asignar de manera inequívoca un espectro a un grupo específico. A la vez, debe evitarse el solapamiento entre los grupos, para no tener ambigüedades entre productos o falsos positivos.

Para calcular la biblioteca principal se usó el algoritmo de distancia Euclidiana en el espacio espectral. La elección del pretratamiento, el rango escogido y el criterio de aceptación (umbral) se realizó de manera que permitiera identificar inequívocamente tantos productos como fuera posible. Para todos aquellos grupos de materias primas que presentaron ambigüedades a partir de la identificación inicial, se crearon sub-bibliotecas de nivel 2 para cada conjunto. Para el cálculo de todas las sub-bibliotecas se empleó el PCA. La optimización de los parámetros (rango, pretratamiento y umbral) para cada modelo se realizó de manera que permitiera diferenciar tanto como fuera posible todas las sustancias que formaban parte de ella. Si dentro de un nivel 2, se continuaban confundiendo productos, se creaba una sub-sub-biblioteca de nivel 3 y, en este trabajo, usando el mismo algoritmo de cálculo y optimizando la combinación de parámetros. Este proceso se repitió hasta que se obtuvo una biblioteca en cascada de n niveles capaz de identificar de manera inequívoca cada material de partida.

4.2.3.3. Validación externa

La validación externa se llevó a cabo con los lotes no utilizados para el cálculo y la optimización de la biblioteca con una excepción. Aquellos productos que en el momento del desarrollo solo tenían un lote, este se usó para la calibración y validación pero los espectros usados en cada conjunto eran distintos. Se evaluó la especificidad y, en aquellos casos en los que había suficientes lotes, la robustez. En los productos que no se pudo evaluar la robustez en el momento de validar la biblioteca, se realizó a posteriori. Hasta que no quedara demostrada la robustez con un total de tres lotes nuevos, el material de partida no se podía identificar únicamente por NIR, sino que se debía realizar con otro método de referencia. Si la robustez fallaba, la nueva variabilidad se incorporaría en una nueva versión de la biblioteca.

La comprobación de la especificidad se realizó mediante la selectividad de la biblioteca. El software de OPUS permite diferenciar 4 situaciones, esquematizadas en la Figura 4.1. Cuando la selectividad es menor a 1 significa que los grupos implicados están solapados. Se pueden dar dos situaciones, que hayan muestras que aparecen en la zona de solapamientos o que ninguna muestra se encuentre en ella. Cuando la selectividad es igual o mayor a 1 significa los grupos están separados. Cuanto mayor es el valor de S , mayor es la distancia entre los grupos. Por lo tanto, el objetivo es obtener una biblioteca donde la selectividad de todos materiales de partida sea mayor a uno, asegurando que no haya ningún falso positivo. Si en algún caso se obtiene una $S < 1$, solo se considera satisfactoria si se da la situación que ningún espectro se encuentra dentro de la zona de solapamiento, y se demuestra con el conjunto de validación externa que en ningún caso se obtendrán falsos positivos.

La robustez se comprobó retando al modelo con lotes nuevos que presentaran la variabilidad esperada para cada producto. El objetivo es que la biblioteca sea capaz de identificarlos correctamente, minimizando o eliminando los falsos negativos y así evitar tanto como sea posible el riesgo de rechazar lotes que son correctos.

4.2.4. Desarrollo de bibliotecas Raman

Dado que las bandas del Raman están bien resueltas, el desarrollo de una biblioteca espectral para la identificación y cualificación de productos es mucho más sencilla. Además el software presenta menos prestaciones por lo que la flexibilidad u opciones para el desarrollo son muy limitadas.

4.2.4.1. Calibración

El proceso empezó con la adquisición de los espectros que se usaron como referencia y sirvieron para crear los métodos de los materiales de partida que debían formar parte de la biblioteca. Este

registro se puede realizar en primera o segunda derivada, lo que significa que a la vez que se registra el espectro de referencia, el software incorporado en el instrumento lo almacena una vez ya se ha calculado la derivada del mismo para ser usados en la identificación. Como norma general, se usó la primera derivada. Una vez registrados, se crearon los métodos correspondientes a los productos definidos. El número de espectros que se emplean para crear un método depende de la variabilidad esperada de la muestra. Se suele usar un espectro de calibración si la variabilidad no afecta a la señal Raman y, en caso contrario, se añaden tantos espectros como sea necesario para cubrirla.

El método estadístico que se usa para la verificación se basa en la probabilidad de que las diferencias observadas entre un nuevo espectro y el espectro de referencia sean causadas por la incertidumbre de la medida. Esta incertidumbre, que está relacionada con la repetitividad de la medida, la calcula automáticamente el software del instrumento mientras se está registrando el espectro de referencia. El valor de este parámetro se basa en la variabilidad esperada en una nueva adquisición debido a las características de la muestra, las propiedades del instrumento y las condiciones ambientales. Esta incertidumbre define el dominio multivariable del espectro de referencia. Al analizar un nuevo producto, este se compara con los espectros de referencia incluidos en el método, si el valor ρ obtenido es igual o mayor a cierto valor de probabilidad (se puede elegir entre 0,05 o 0,1) se acepta la hipótesis nula (el espectro de la nueva muestra pertenece a la misma población que el espectro de referencia)¹⁴.

4.2.4.2. Validación

El software permite realizar un test de selectividad para demostrar que no hay confusiones entre productos. Esto es, para un producto concreto, solo se obtiene un valor de ρ mayor al valor de aceptación determinado para un solo método, y no para las otras clases incluidas en la biblioteca Raman. La robustez se demuestra con la correcta identificación de tres lotes no incluidos en la calibración de la biblioteca.

4.3. RESULTADOS Y DISCUSIÓN

4.3.1. Biblioteca NIR

4.3.1.1. Evaluación de la biblioteca anterior

La estructura de la biblioteca de identificación NIR de la versión anterior constaba de una biblioteca principal (nivel 1) con 184 materiales de partida, 15 sub-bibliotecas de nivel 2 y 5 de nivel 3.

La evaluación previa que se realizó de la biblioteca de identificación NIR, llevó a detectar que de 184 materiales de partida contenidos en el método, había 80 productos que se identificaban parcial

o totalmente de manera incorrecta. Se realizó un estudio que permitiera determinar los motivos de esta baja robustez, y así establecer la estrategia que se seguiría para actualizar la biblioteca. No había un motivo único que provocara los errores en la identificación, sino que era un conjunto de causas.

Incluir toda la variabilidad esperada en un producto es un requisito imprescindible para el desarrollo de un método NIR. Si en el momento de realizar la biblioteca no se ha incluido toda la diversidad que presenta un producto ya sea debido a: (1) que no se dispone de lotes suficientes, (2) que aunque se dispone de una cantidad elevada de lotes, estos no representan la variabilidad total esperada o (3) que con el tiempo, el producto ha evolucionado; la biblioteca de identificación NIR no tendrá la robustez necesaria. El alcohol cetoesarílico es un ejemplo de esta situación (Figura 4.3). En este caso, hubo una serie de lotes que se identificaron correctamente y otros que no. Los espectros no identificados presentan una línea base diferente a los que estaban incluidos en el conjunto de calibración que no se corregía con el pretratamiento aplicado en el modelo. Esto es un indicativo de que hay nueva variabilidad que no se detectó o no se tuvo en cuenta al realizar la biblioteca. En todos los casos en los que se daba esta situación, fue necesario incorporar la nueva variabilidad dentro del conjunto de calibración para la actualización de la biblioteca NIR.

La manera de adquirir los espectros influye de forma directa a los espectros. Para el desarrollo de la versión anterior, los espectros de los productos líquidos se registraron con la sonda de transmisión. Debido a una serie de problemas técnicos, se tomó la decisión que, a partir de cierto momento, los espectros de las sustancias semisólidas y líquidas se adquirieran usando el compartimento de transmisión mediante cubetas. Como era de esperar, la diferencia de intensidades que presentan los espectros adquiridos de ambas formas (Figura 4.4) imposibilitó la correcta identificación de lotes registrados de la nueva manera. Además, esta diferencia no permitía que se pudieran incluir los espectros adquiridos mediante cubeta y mantener los anteriores (con sonda). Esto obligó a eliminar

del conjunto de calibración los registrados con la sonda y considerar únicamente los registrados con cubeta. Incluso en algunos casos fue imposible mantener el producto en la misma biblioteca, ya que el rango espectral que se usa en el modelo de la biblioteca principal (que es necesario usarlo porque ha demostrado suficiente poder de identificación) el instrumento presenta una saturación del detector, lo que se traduce al espectro en forma de ruido elevado. En la Figura 4.5 se muestra un ejemplo donde se observa el fenómeno, con el excipiente Emulsión de Simeticona. Aunque se incluyeran todos los espectros en la calibración, nunca una nueva muestra se identificará correctamente debido a la aleatoriedad del ruido. En esta situación, se decidió que era imprescindible desarrollar una segunda biblioteca que incluyera todos los productos que presentaban esta singularidad. Esta segunda biblioteca no forma parte de esta tesis doctoral.

En el caso de los productos sólidos que se suministran con bolsas de polietileno transparentes como envase primario, el registro de los espectros se realiza a través de ellas con la sonda de reflectancia difusa. Aunque es una ventaja desde el punto de vista operativo, es un reto a la hora de desarrollar la biblioteca. El plástico puede generar alteraciones en el espectro debido a la señal del mismo¹⁵. Al realizar la comparación de un API registrado directamente sobre el producto y a través del plástico (Figura 4.6), se observan claramente bandas que pertenecen al polietileno. Para minimizar la aparición de esta señal en los espectros de los productos, el blanco se registra a través de la misma bolsa de plástico. Ahora bien, por causas operacionales como son pequeñas diferencias del grosor del plástico o lecturas a través de pequeñas arrugas en la bolsa, la señal del polietileno es inevitable. Por lo que se debe aceptar cierto grado de variabilidad en esa zona del espectro, pero evitando la incorporación de espectros cuando la variabilidad empieza a ser significativa, ya que podría enmascarar señales importantes de nuevas muestras.

4.3.1.2. Desarrollo de la nueva biblioteca de identificación NIR

La biblioteca principal se desarrolló de tal forma que sirviera para identificar de manera inequívoca tantos productos como fueran posibles. El algoritmo escogido para el cálculo fue la distancia Euclidiana en el espacio espectral. Se usó como pretratamiento la primera derivada con una ventana móvil de 17 puntos seguida de una SNV. El rango espectral seleccionado fue 9000-4400 cm^{-1} . Se descartaron los rangos 12500-9000 cm^{-1} y 4400-4000 cm^{-1} debido a que no presentaban información útil que pudieran contribuir a una mejora sustancial de la identificación. El valor de aceptación o distancia umbral para cada material de partida se ajustó a un nivel de confianza del 97%. Los espectros de nuevas muestras quedarían identificados como pertenecientes a un producto determinado si la distancia calculada era menor que el umbral definido para un grupo. Como resultado, la biblioteca principal permitía identificar de manera inequívoca 129 materiales de partida de los 224 incluidos. Entre ellos, la mayoría de los de los APIs. Solo dos principios activos se confunden con un grupo de excipientes. Un total de 95 sustancias presentaban

ambigüedades, por lo que fue necesario crear 20 sub-bibliotecas en el segundo nivel y, en algunos casos, se tuvieron que incluir varios niveles extras para acabar resolviéndolas. Entre los 20 grupos se encontraban polisacáridos (almidones y celulosas por separado), cadenas hidrocarbonadas largas, sacarosas, lactosas, lacas, etc. Los componentes denominados lacas son, en realidad, mezclas con sustancias de carga mayoritarias y comunes, lo cual explica la dificultad de su diferenciación mediante espectroscopia NIR.

El objetivo de desarrollar sub-bibliotecas es definir un modelo que permita cualificar de manera inequívoca productos químicamente similares. Debido a que los espectros son parecidos, el algoritmo que se usó fue también la distancia Euclidiana, pero una vez realizada la factorización. Esto es, se calculó la distancia entre los *scores* una vez reducida la dimensionalidad de los datos mediante el PCA. Esto se hace, precisamente, porque los datos originales están altamente correlacionados. El rango espectral, el pretratamiento de los datos, el número de PCs y el umbral para cada grupo se optimizaron de manera que permitiera identificar/cualificar correctamente todas

las materias primas sin que las delimitaciones de los grupos se solaparan. En algún caso no fue posible resolver las diferencias entre productos, por lo que se debieron agrupar productos químicamente iguales pero que presentaban distintas propiedades físicas. Un ejemplo de esto se dio con un tipo de microcelulosa que, más adelante, se comentará con detalle. La estructura final de la biblioteca completa está esquematizada en la Figura 4.7. En total está compuesta por 5 capas (niveles) los cuales sirven para realizar una identificación en cascada. La biblioteca principal, como ya se ha comentado, consta de 224 productos. En el segundo, tercer, cuarto y quinto nivel hay un total de 95, 38, 15 y 4 sustancias respectivamente.

A continuación se describen en detalle algunas de las ramas de las cascadas que han sido necesarias desarrollar para cualificar los materiales no identificados en el primer nivel o biblioteca principal.

Almidones

Cuando se analiza el gráfico de *scores* (Figura 4.8) que se ha obtenido con el modelo definido para la sub-biblioteca Nivel 2 grupo A de la Figura 4.7, se detecta como los almidones y los productos derivados de estos forman un clúster. Para resolver el solapamiento que presentan estas sustancias, es necesario desarrollar la sub-biblioteca Nivel 3 Grupo A de la Figura 4.7. También cabe mencionar que fue necesario desarrollar otra sub-biblioteca (Nivel 3 Grupo B de la Figura 4.7) para poder identificar de manera inequívoca el conjunto de cápsulas de gelatina duras de diversos tamaños y colores y el sulfato de magnesio heptahidratado.

El almidón es un polisacárido que contiene aproximadamente un 75% de amilopectina y un 25% de amilosa. Ambas tienen como cadena principal la α -D-Glucosa con enlaces α -(1 \rightarrow 4). La

MÉTODOS DE IDENTIFICACIÓN

amilopectina, además, presenta ramificaciones de enlaces α -(1 \rightarrow 6)¹⁶. En la Tabla 4.2 se describen todos los materiales de partida incluidos en el grupo sujeto de estudio.

Tabla 4.2 Productos incluidos en la sub-biblioteca Nivel 3 Grupo A, incluidas su funcionalidad y características.

Compuesto	Funcionalidad⁶	Características
Almidón pregelatinizado de maíz	Excipiente para la dilución o disgregación de comprimidos y cápsulas, o como agente aglutinante en comprimidos.	Almidón procesado química o físicamente para eliminar o reducir los gránulos de almidón ⁶ .
Almidón de maíz	Excipientes para la dilución o disgregación de comprimidos y cápsulas,	Almidón que proviene del maíz.
Almidón de arroz	o como agente aglutinante en comprimidos o agente espesante.	Almidón que proviene del arroz
Carboximetilalmidón sódico tipo A	Excipiente para la disgregación de comprimidos y cápsulas.	Es la sal sódica del carboximetil éter del almidón.
Maltodextrina	Excipiente con varias funcionalidades. Agente de recubrimiento, disgregante para comprimidos y cápsulas...	Mezcla de polisacáridos que proviene de la hidrólisis parcial con ácido o enzimas del almidón ⁶ .
Aroma de fresa	Excipiente para dar aroma al producto farmacéutico.	El componente de carga mayoritaria es maltodextrina.
Raíz de altea	API	Las raíces son ricas en almidón.

El modelo calculado usando como pretratamiento la primera derivada de 17 puntos de ventana móvil seguido de una SNV permitió diferenciar claramente entre el grupo de almidones y carboximetil almidón por un lado y el grupo de maltodextrina y aroma de fresa por otro. La raíz de altea se identificó de manera inequívocamente en esta sub-biblioteca. En la Figura 4.9 se observan las diferencias entre los grupos previamente mencionados en el gráfico de *scores* obtenido una vez se ha aplicado el PCA y definidos los umbrales. La diferencia entre los productos que forman los dos clústeres heterogéneos, almidones por un lado y maltodextrinas por el otro, es el número de unidades que componen las moléculas.

Para el grupo de almidones y carboximetil almidón sódico aún fue necesario crear dos niveles más. El primero diferenciaba sin ninguna confusión el almidón pregelatinizado (diferencia en el tamaño de partícula respecto a los demás productos del grupo), y el carboximetil almidón sódico (químicamente distinto a los demás productos que forman parte del modelo), aplicando una primera derivada de 25 puntos seguido de una SNV en el rango espectral 6000-12500 cm^{-1} . Pero aún no era posible cualificar los dos almidones con distinto origen, el de arroz y el de maíz. Introduciendo un quinto nivel para estas dos materias primas, y aplicando un SNV entre 8700 y 9500 cm^{-1} fue suficiente para cualificar de manera correcta ambos productos. En la Figura 4.10 se pueden observar los espectros y el gráfico de *scores* de este último nivel (Nivel 5 Grupo A).

Por último, la otra bifurcación de la sub-biblioteca que se creó para el grupo de maltodextrinas (Nivel 4 Grupo A), se aplicó una primera derivada de 25 puntos en el rango 7000-10000 cm^{-1} y el modelo permitió identificar ambos productos, la maltodextrina por un lado, y el aroma de fresa por el otro.

Celulosa microcristalina

En el Nivel 2 Grupo C de la Figura 4.7 de la biblioteca de identificación por NIR se agruparon todas las celulosas microcristalinas y derivadas de estas. La celulosa microcristalina (MCC) es una celulosa purificada y parcialmente despolimerizada. Se prepara mediante el tratamiento con ácidos minerales de la α -celulosa que se obtiene de la pulpa de fibras de plantas¹⁷. La celulosa es un polisacárido formado por D-glucosa con enlaces β -1,4-O-glucosídicos. En la Tabla 4.3 se resumen todos los productos que conforman el modelo junto a sus funcionalidades y características.

Tabla 4.3 Productos incluidos en la sub-biblioteca Nivel 2 Grupo C, incluidas su funcionalidad y características.

Compuesto	Funcionalidad ⁶	Características
Celulosa microcristalina	Absorbente, agente de	Celulosa fabricada por la compañía A

Tabla 4.3 Productos incluidos en la sub-biblioteca Nivel 2 Grupo C, incluidas su funcionalidad y características.

Compuesto	Funcionalidad⁶	Características
65µm Fabricante A	suspensión, diluyente en cápsulas y comprimidos, disgregante en comprimidos	con tamaño de partícula 65µm
Celulosa microcristalina 130µm Fabricante A		Celulosa fabricada por la compañía A con tamaño de partícula 130µm
Celulosa microcristalina 50µm Fabricante B		Celulosa fabricada por la compañía B con tamaño de partícula 50µm
Celulosa microcristalina 100µm Fabricante B		Celulosa fabricada por la compañía B con tamaño de partícula 90µm
Celulosa microcristalina silificada HD 90	Diluyente de cápsulas y comprimidos	Mezcla de celulosa microcristalina con aproximadamente un 2% en peso de dióxido de silicio coloidal.
Celulosa microcristalina y carmelosa sódica	Agente dispersante, estabilizador de emulsiones, agente estabilizador, agente de suspensión y espesante.	Contiene aproximadamente un 11% de carmelosa sódica.

Con la sub-biblioteca Nivel 2 Grupo C se pudo diferenciar de forma clara las dos celulosas microcristalinas del proveedor A, las dos del proveedor B, la silificada y finalmente la que contiene carmelosa sódica. Para ello se realizó un pretratamiento de los datos aplicando una segunda derivada de 13 puntos de ventana móvil seguido de una SNV en el rango espectral 7500-6400 cm^{-1} . Los componentes principales seleccionados para definir el modelo fueron el segundo y el cuarto, ya que son los que permitían una mayor diferenciación entre los grupos mencionados anteriormente. Las diferencias espectrales que presentan los productos se traducen en los clústeres formados en el gráfico de *scores* (Figura 4.11) del PCA calculado. Este primer modelo permite identificar correctamente los productos químicamente distintos (contienen otras sustancias aparte de MCC), y dentro de las MCC puras, diferenciar entre los dos fabricantes homologados de la materia prima. Ahora bien, no es capaz de cualificar entre MCC de distinto tamaño de partícula. En este punto, sería necesario incorporar dos nuevos niveles (uno para cada fabricante) para poder cualificar correctamente las MCC.

Como se puede comprobar al observar el esquema general de la biblioteca de la Figura 4.7, finalmente solo se incluyó una nueva sub-biblioteca (Nivel 3 Grupo F), la cual permitía diferenciar las MCC que provenían del fabricante B por su tamaño de partícula. Aplicando una primera derivada de 13 puntos seguido de una SNV en los rangos 7500-5900 y 5400-5000 cm^{-1} , se observan diferencias (Figura 4.12). Ello es debido a la diferencia de *scattering* que tiene lugar en productos químicamente equivalentes pero con una diferencia en el tamaño de partícula (en este caso 50 y 100 μm). Estas variaciones permitieron desarrollar un modelo con suficiente poder discriminatorio entre ambos materiales.

Por otro lado, no fue posible encontrar una combinación óptima que permitiera diferenciar por tamaño de partícula las MCC del fabricante A. Esto fue sorprendente, ya que la diferencia de tamaños de partícula de los materiales son aproximadamente equivalentes que los del fabricante B (en este caso 65 y 130 μm). Para encontrar una explicación a este hecho, inicialmente se evaluó la distribución del tamaño de partícula y la variabilidad que presentaban distintos lotes de los productos de ambos proveedores. Este estudio comparativo se realizó con un equipo de difracción láser, el cual permite determinar la distribución del tamaño de partícula de un material mediante tres parámetros: d_{10} , d_{50} y d_{90} . Estos valores están basados en la distribución del volumen de las partículas. En la Figura 4.13 se puede observar como las distribuciones de tamaño de partícula de las dos clases de MCC para cada fabricante son equivalentes, y presentan una variabilidad parecida entre lotes. Este hecho llevó a descartar que se tratara de un problema del producto. La diferencia entre ambos fabricantes es el número de bultos que se reciben en cada lote. Para el fabricante A, cada lote consta de aproximadamente 300 sacos, mientras que para el B este número es menor (20-40). No es lo mismo registrar 300 espectros de manera consecutiva que 20. Se concluyó que por causas operacionales, el método analítico no ofrece el poder resolutivo suficiente que permita cualificar las dos MCC del fabricante A según su tamaño de partícula. Se decidió incorporar en una

misma clase ambos productos, generando un grupo de “celulosa microcristalina fabricante A”. El modelo comentado anteriormente, el del Nivel 2 Grupo C, permitía identificar esta clase de manera inequívoca, sin necesidad de añadir una sub-biblioteca. Se postula que, en este caso, cuando un lote de cualquiera de las MCC del fabricante A se haya identificado correctamente como “celulosa microcristalina fabricante A” mediante NIR, en el laboratorio de control de calidad se analizará la muestra representativa de cada lote para verificar que el tamaño de partícula es el correcto.

Lactosas

En el Nivel 2 Grupo G de la Figura 4.7 de la biblioteca de identificación por NIR se agruparon todas las lactosas disponibles. La lactosa es un disacárido formado por una galactosa y una glucosa con un enlace (1→4). Dependiendo de las condiciones de cristalización del excipiente, se pueden obtener cristales α -lactosa, β -lactosa o forma amorfa¹⁸. En la Tabla 4.4 se describen las funcionalidades y características de los 4 productos incluidos en esta rama.

Tabla 4.4 Productos incluidos en la sub-biblioteca Nivel 2 Grupo G, incluidas su funcionalidad y características.

Compuesto	Funcionalidad⁶	Características
Lactosa monohidratada	Excipiente para la dilución de comprimidos y cápsulas, o como agente aglutinante en comprimidos.	α -lactosa con tamaño de partícula de $\sim 100\mu\text{m}$
Lactosa monohidratada granulada		α -lactosa con tamaño de partícula de $\sim 180\mu\text{m}$
Lactosa monohidratada polvo		α -lactosa con el 60-85% de partículas con un tamaño menor a $75\mu\text{m}$

Tabla 4.4 Productos incluidos en la sub-biblioteca Nivel 2 Grupo G, incluidas su funcionalidad y características.

Compuesto	Funcionalidad⁶	Características
Lactosa atomizada	Excipiente para la compresión directa de comprimidos, diluyente de comprimidos y cápsulas	α -lactosa mezclada con la forma amorfa con tamaño de partícula de $100\mu\text{m}$

La sub-biblioteca Nivel 2 Grupo G sirvió para realzar la diferencia entre la lactosa que contiene una mezcla de forma cristalina α y la forma amorfa del producto del resto de lactosas monohidratadas, que son α -lactosas. Para ello se aplicó una primera derivada de 9 puntos seguido de la SNV en el rango espectral de $7900\text{-}5900\text{ cm}^{-1}$ y se seleccionó el segundo componente principal. La utilidad del NIR para diferenciar formas cristalinas y amorfas ha sido altamente descrita en la bibliografía. En la Figura 4.14 se observan las diferencias espectrales entre las dos formas incluidas en el modelo.

La incorporación de un tercer y cuarto nivel (Nivel 3 Grupo H y Nivel 4 Grupo E respectivamente) fueron necesarios para acabar de resolver las ambigüedades que presentaban las lactosas monohidratadas y poderlas cualificar de manera inequívoca según su tamaño de partícula.

4.3.1.3. Validación de la biblioteca NIR

Una vez definidos y optimizados todos los modelos que componen la biblioteca de identificación NIR, se realizó la validación de dicha biblioteca para comprobar su selectividad y robustez. En el histograma de selectividad obtenido mediante la validación de la biblioteca (Figura 4.15) se representa el valor de S. Esta se usa para comparar todos los grupos entre ellos y detectar si hay solapamientos. El cálculo de S se realiza mediante la división entre la distancia de los espectros promedios de cada grupo y la suma de los umbrales asignados a cada grupo⁵. Aunque el objetivo principal es obtener todos los valores de S igual o mayor a 1, una cantidad mínima de grupos se solapaban. Los valores de S eran mayores a 0,8 y en ningún caso había espectros que cayeran en la zona de solapamiento entre los grupos. Después de evaluar los riesgos y comprobar mediante el

análisis de nuevos lotes que en ningún caso se obtendrían falsos positivos se concluyó que esta era suficientemente selectiva. Aunque se haya demostrado que la en ningún caso habrá confusiones entre los productos, la muestra representativa de estos materiales de partida se identificarán por el método alternativo en el laboratorio de control de calidad.

Todos los lotes que no fueron usados para la calibración y optimización de la biblioteca se usaron para validar la robustez de la misma. En todos los casos se obtuvieron identificaciones correctas sin ningún falso positivo ni falso negativo. Por lo que se dio como demostrada la robustez de la biblioteca. En aquellos casos que el número de lotes disponibles fue menor a 3, no se dio como demostrada la robustez. Hasta que no se identificaran correctamente tres lotes nuevos, el producto no se podría identificar únicamente por NIR, sino que sería necesario usar otro método alternativo.

4.3.2. Biblioteca Raman

La espectroscopia Raman se trata de una técnica completamente nueva en la compañía que servirá para desarrollar una biblioteca para la verificación de materiales de partida. Cuando se utiliza una biblioteca espectral para la identificación o cualificación de un producto, esta puede emplearse de dos formas: comparar el nuevo espectro con todas las clases incluidos en la biblioteca (identificación) o sabiendo previamente a que clase pertenece, escogerla y que el nuevo espectro se compare únicamente con dicho grupo (verificación). Antes de definir el protocolo para el desarrollo de una biblioteca Raman fue necesario tener un conocimiento de lo que era capaz de identificar y lo que no. En la Tabla 4.5 se enumeran posibles parámetros de análisis o características de los productos que podrían tener un impacto en la señal Raman, y por consiguiente, en la identificación de un producto.

Tabla 4.5 Relación de los parámetros de análisis o características de la muestra que pueden tener un efecto en la señal Raman.

Parámetro o característica	Posible efecto en la señal
Fluorescencia	Se presenta como una línea base que puede llegar a ser tan elevada que enmascare las bandas Raman del producto. La muestra se puede degradar

Tabla 4.5 Relación de los parámetros de análisis o características de la muestra que pueden tener un efecto en la señal Raman.

Parámetro o característica	Posible efecto en la señal
	térmicamente por la absorción de la radiación.
Color de la muestra	La muestra se puede degradar térmicamente por la absorción de la radiación.
Polimorfismo	Distintas formas cristalinas y amorfas pueden presentar diferencias en las bandas.
Tamaño de partícula	Algunos autores apuntan que la señal Raman es sensible al tamaño de partícula debido a la presencia de scattering ¹³ .
Impurezas	Hay impurezas que pueden provocar fluorescencia.
Isómeros	Diferencia en las bandas según el isómero que se analiza.
Componentes minoritarios	No detectar la presencia de componentes minoritarios.
Concentración o riqueza	La diferencia en la intensidad de ciertas bandas del espectro puede afectar a la correcta identificación.
Envase primario	El espectro Raman puede presentar bandas que correspondan al envase primario.
Temperatura ambiente	Puede afectar: <ul style="list-style-type: none"> - Al correcto funcionamiento del equipo (detector y láser). - A la intensidad de las bandas, ya que son función de la temperatura. Afecta de manera distinta a las bandas de alta y baja frecuencia¹³.
Fabricante	Distintos espectros debido a diferentes propiedades de los productos.

No fue necesario realizar un estudio en profundidad de todas las variables de la Tabla 4.5, ya que se pudieron descartar algunas. La primera de ellas fue la temperatura ambiente. El análisis se realizará en una sala de muestreo donde la temperatura y la presión están completamente controladas, y se mueven en un rango muy estrecho. Por lo que no debería producirse ninguna variación en el espectro debido a este motivo. El segundo fue el color de la muestra. El instrumento tiene un mecanismo que no permite realizar medidas prolongadas, cuando detecta que el tiempo empleado para registrar el espectro es demasiado largo, el láser se para y en la pantalla avisa de que se cambie de posición la muestra. Excepto los dos casos mencionados, el resto de características y variables se evaluaron con muestras reales.

4.3.2.1. Fluorescencia

La fluorescencia es una de las desventajas más importantes que presenta la técnica Raman. Al producirse este fenómeno, aparece una señal de fondo en el espectro, pudiendo ser tan elevado que llegue a enmascarar las bandas del Raman. Aunque el uso de un láser de 785nm presenta un buen compromiso entre la sensibilidad y fluorescencia, está documentado que hay algunos excipientes que presentan fluorescencia, como por ejemplo la celulosa microcristalina (MCC) o la croscarmelosa sódica¹⁹.

Al obtener el espectro del MCC de dos fabricantes distintos, aunque presentaban una señal de fondo causada por la fluorescencia, se observaban bandas Raman. Al pasar a analizar el producto MCC con un 11% de carmelosa sódica, las bandas quedaron enmascaradas casi en su totalidad. Al

calcular los espectros en primera derivada, se confirmó que la señal Raman registrada con las MCC era suficiente como para poder identificar dicho producto. Por lo contrario, la relación señal ruido obtenida en el espectro del producto MCC con carmelosa sódica era tan baja que no se podía emplear dicha técnica para verificar el producto por Raman (Figura 4.16).

4.3.2.2. Polimorfismo

Siendo el Raman una técnica conocida para poder diferenciar entre formas cristalinas, se compararon los espectros de dos tipos de lactosas, la que tiene forma cristalina α y la que es una mezcla de forma cristalina α y amorfa. La comparación de los espectros primarios demostraron un desplazamiento en la línea de base, la cual era reducida casi completamente una vez aplicada la primera derivada (Figura 4.17). No había ninguna señal distinta entre la forma α -lactosa y la amorfa, por lo que en este caso no se podrían cualificar mediante Raman.

Por otro lado, se evaluó si la presencia o ausencia de polimorfos podría ser detectado o caracterizado por el instrumento Raman. Se creó un método con el espectro de referencia de un API que no contenía el polimorfo B. Posteriormente se analizó el mismo API con un 10 y un 20% de polimorfo B. La Tabla 4.6 confirma que ambas muestras se identificaron correctamente ($p \geq 0.05$). Se concluyó el equipo utilizado no es sensible a diferentes grados de polimorfismo para este API.

Tabla 4.6 Resultados verificación API con distintos grados de polimorfo B.

Muestra	Polimorfo B (% en peso)	Valor ρ	Resultado
Muestra 02	10	0.418	Pasa
Muestra 04	20	0.346	Pasa

4.3.2.3. Tamaño de partícula

Los mismos ejemplos expuestos en la Figura 4.16 y en la Figura 4.17 sirvieron para observar las diferencias que presentaban los espectros Raman debido al tamaño de partícula. Aunque en los datos primarios, se observa una diferencia en la línea base de los productos con distinto tamaño de partícula (MCC de 50 y 90 μm por un lado, y lactosa monohidratada y lactosa monohidratada polvo por otro). Esta diferencia es completamente reducida cuando se aplica la primera derivada. Si bien con la comparación visual de los espectros parece evidente que no se diferenciarán cuando se analice por Raman, se creó un método que permitiera obtener la evidencia numérica. Para ello se usó como espectro de referencia la MCC 50 μm , y se verificaron dos nuevos lotes, uno de MCC 50 μm y otro de 90 μm . Como era de esperar, ambos se identificaron como MCC 50 μm (Tabla 4.7).

Tabla 4.7 Resultados verificación MCC con distintos tamaño de partícula.

Muestra	Valor ρ	Resultado
MCC 50 μm	0.176	Pasa
MCC 90 μm	0.539	Pasa

4.3.2.4. Impurezas

La presencia de ciertas impurezas puede provocar fluorescencia, ya que es un proceso más eficiente que la dispersión Raman²⁰. La sensibilidad del instrumento se comprobó con un API que presentaba distintos perfiles de impurezas. Se desarrolló el método de verificación con los espectros de dos lotes. Al analizar tres lotes nuevos del mismo API, dos lotes se identificaron correctamente obteniendo valores ρ superiores a 0.05. para el tercer lote, el resultado de la identificación fue “Fallo” obteniendo un valor ρ de 0.00. Al comprobar el espectro de este último

lote y compáralo con el resto, se detectó la presencia de una señal de fondo, la cual era inexistente en los espectros usados en la calibración. Se evaluó el perfil de impurezas de todos los lotes analizados por Raman, tanto los usados para la calibración, como los tres nuevos. En la Tabla 4.8 se puede ver que el lote incorrecto tiene un alto contenido de impureza C, si se compara con los demás lotes, aunque se encuentra dentro de los límites especificados. Se atribuyó que dicha impureza era la responsable del fenómeno fluorescencia.

Tabla 4.8 Perfil de impurezas de distintos lotes de un API con los valores de ρ obtenidos en el análisis por Raman de dicho API.

	Lote 1 API referencia	Lote 2 API referencia	Lote 3 API analizado	Lote 4 API analizado	Lote 5 API analizado
Impureza B	NQ ^a	ND ^b	0.055%	ND ^b	0.060%
Impureza C	NQ ^a	0.052%	0.057%	0.236%	ND ^b
Impureza D	NQ ^a	NQ ^a	0.089%	0.097%	0.080%
Impureza E	NQ ^a	NQ ^a	0.070%	ND ^b	0.092%
Impureza F	0.114%	0.081%	0.109%	0.116%	0.167%
Valor ρ	NA ^c	NA ^c	0.6382 (Pasa)	0.0000 (Fallo)	0.6306 (Pasa)

^aNo cuantificable; ^bNo detectable; ^cNo aplica

4.3.2.5. Isómeros

La espectroscopia Raman es una técnica que permite diferenciar entre isómeros¹³ ya sea porque hay distintos modos vibracionales activos al Raman o porque presentan una red cristalina distinta. Se evaluó la capacidad del instrumento a diferenciar los isómeros de dos polialcoholes que se usan como edulcorantes en la industria. Estos fueron el manitol ((2R,3R,4R,5R)-hexano-1,2,3,4,5,6-hexol) y el sorbitol ((2S,3R,4R,5R)-hexano-1,2,3,4,5,6-hexol), que se diferencian por la orientación del grupo hidroxilo en el segundo carbono, son estereoisómeros. Al comparar los espectros

obtenidos, Figura 4.18, se observan bandas distintas en cada uno de los compuestos, por lo que se confirma que es posible identificar de manera inequívoca compuestos isómeros.

4.3.2.6. Componentes minoritarios

El límite de detección (LOD) de la espectroscopia Raman es difícil de establecer, ya que depende de la señal Raman que presenta el producto objetivo, así como de la señal que presentan los otros compuestos que forman parte de la muestra. Una buena aproximación es considerar un LOD entre un 1 y un 2%¹⁹. Aunque la teoría es esta, era necesario comprobar la sensibilidad del instrumento para diferenciar productos iguales, pero con distinto contenido de componentes minoritarios. Para este fin, se registró el espectro de una sacarosa polvo y de una sacarosa polvo con un 2% de contenido de sílice. En la Figura 4.19 no aparece ninguna diferencia entre los espectros, por lo que en este caso no es posible diferenciar ambos productos. Esta conclusión no se puede generalizar, ya que la aparición de señales del componente minoritario dependerá de la efectividad de la dispersión Raman del mismo. Por lo tanto, deberá estudiarse cada caso de forma individual.

4.3.2.7. Concentración

Se comprobó como podía afectar la riqueza o concentración presente en un producto para llegar a ser identificado correctamente. Para ello se creó un método con el espectro de referencia de un producto granulado que contenía un 11.7% de API. Después de comprobar que las señales del espectro Raman eran intensas, sin ningún componente que generara fluorescencia, se analizaron dos lotes nuevos del producto. En este caso, contenían un 17.5% y un 21.7% de API. Como se demuestra en la Tabla 4.9 ambos productos se identificaron correctamente. Se puede concluir que la concentración, y en extensión la riqueza, no afecta en la identificación de sustancias.

Tabla 4.9 Resultados obtenidos en la verificación de un producto granulado con distintas concentraciones de API

Nombre del producto	Concentración (% en peso)	Valor p	Resultado verificación
Producto granulado A	17.5	0.515	Pasa
Producto granulado B	21.7	0.552	Pasa

4.3.2.8. Envase primario

El uso de un Raman portátil tiene que servir para poder minimizar el tiempo empleado en la verificación de materiales de partida. El 80% de productos se reciben en bolsas de polietileno como envase primario. Se ha evaluado la contribución del polietileno en el espectro. Para ello se registró un espectro de un API directamente sobre el producto y otro a través de la bolsa de polietileno. Al comparar los espectros obtenidos, Figura 4.20, no se detectó ninguna diferencia entre ambos, por lo que no hay señal procedente del polietileno. Esto es debido a que el punto focal del láser se encuentra más allá de la bolsa, en la muestra. Es posible registrar los espectros a través de los envases primarios de este material. Se deberá tener especial atención en aquellos casos donde la señal del Raman sea pobre, ya que en este caso sí podría haber contribuciones del envase primario en el espectro.

4.3.2.9. Fabricante

En algunos casos los materiales de partida están homologados para varios fabricantes. Es importante entender si las posibles diferencias entre ellos podrían afectar a la identificación. Se realizó la comparación de un API con dos fabricantes autorizados y sus espectros se comparan en la Figura 4.21. La fluorescencia que presenta el API del fabricante A influye de tal manera que no es posible crear el método de identificación para ambos fabricantes con un solo espectro de referencia, será necesario incluir uno de cada uno. Esto podría ser debido a la presencia de ciertas impurezas que provocan fluorescencia.

4.3.2.10. Indicaciones para el desarrollo de una biblioteca Raman

Las conclusiones de los resultados obtenidos en este primer estudio, se pueden resumir en la Tabla 4.10. Como es evidente, esta es una primera aproximación para entender las capacidades del instrumento. Cuando se desarrolla una biblioteca, en caso de duda, se realiza una comparación o comprobación previa, para determinar los efectos antes de desarrollar el método.

Tabla 4.10 Relación de los parámetros de análisis o características de la muestra que pueden tener un efecto en la señal Raman.

Parámetro o característica	Mitigación del efecto
Fluorescencia	Comprobar si hay señal Raman. Si no se observan bandas -la relación señal ruido es baja- entonces se debe concluir que el Raman no es una técnica para verificar dicha sustancia.
Color de la muestra	Obtener el espectro de referencia en una muestra representativa del lote. Evitar el registro en el envase original por si falla el sistema automático de paro de registro y así evitar degradación de la muestra. No hay problema en el análisis en rutina, ya que la adquisición dura segundos.
Polimorfismo	Evaluar cada caso en particular. Registrar un espectro de referencia para cada polimorfo. En principio no se observarán diferencias entre los espectros. Se debe crear un único método que agrupe todas las formas cristalinas y amorfas.
Tamaño de partícula	No afecta en la identificación de compuestos. Crear un único método con un espectro de referencia de un tamaño de partícula en concreto. Demostrar en la validación la correcta identificación de productos con distinto tamaño.
Impurezas	Pueden afectar. Adquirir espectros con distintos perfiles de impurezas y que se encuentren dentro de especificaciones e incluirlos como espectros de referencia.
Isómeros	Pueden presentar espectros distintos debido a la presencia de distintos modos vibracionales activos o por diferencias en la red cristalina. Crear un método para cada uno.

Tabla 4.10 Relación de los parámetros de análisis o características de la muestra que pueden tener un efecto en la señal Raman.

Parámetro o característica	Mitigación del efecto
Concentración	Depende de la efectividad de dispersión Raman del componente minoritario. Si no afecta, el registro de un solo espectro de referencia para crear un solo método será suficiente.
Envase primario	Si es bolsa de polietileno no afecta. Por lo que se puede analizar a través de ella. Si la señal de Raman es débil, evaluar si aparecen bandas correspondientes al envase. En el caso de tener otros envases primarios, como son cartón, aluminio, plástico opaco... los espectros se registrarán sobre el producto.
Temperatura ambiente	No afecta.
Fabricante	Los productos provenientes de más de un fabricantes pueden presentar distintas propiedades y estas pueden hacer que los espectros presenten ciertas diferencias, por lo que se incluirá un espectro de referencia de cada fabricante para la creación del método.

4.4. CONCLUSIONES

Se ha llevado a cabo una actualización de la biblioteca de identificación de materia prima mediante NIR ya existente. Esta permite la identificación y cualificación de los materiales de partida en la cabina de muestreo que se encuentra en del almacén de entrada, por lo que se trata de un análisis *at-line*. Esta nueva versión ha servido para incorporar 44 productos nuevos y aumentar la robustez de los materiales ya incluidos en la versión antigua. Con los resultados obtenidos, se verifica la capacidad de la técnica de análisis NIR de cualificar materiales según su tamaño de partícula, forma cristalina y longitud de cadenas poliméricas.

Por necesidades industriales se ha adquirido un instrumento Raman para la identificación de materia prima. Al tratarse de una técnica nueva en la empresa, se ha realizado un estudio previo de posibles efectos en la señal Raman que ha servido como base para establecer el protocolo del desarrollo de una biblioteca. Se ha determinado que la presencia de impurezas, la fluorescencia que presenta algún producto y el fabricante, son efectos que deben estudiarse antes de la creación de un método. Por otro lado, parece que el polimorfismo, el tamaño de partícula, la presencia de componentes minoritarios o la riqueza no afectan en la identificación del producto. También se ha confirmado la habilidad de la técnica en diferenciar estereoisómeros. En el caso de componentes minoritarios, será necesario realizar un estudio previo, ya que depende de la efectividad de la intensidad de la señal Raman de dicho producto. En el caso del polimorfismo, también se hará una comparativa de los espectros obtenidos de cada forma, para detectar la presencia de señales distintas.

Al comparar las posibilidades del NIR y el Raman para la identificación y cualificación de materiales de partida, parece que se puede concluir que el Raman es una técnica adecuada para la

identificación de API y algunos excipientes. Pero el NIR es la técnica por excelencia para, no solo identificar, sino cualificar excipientes. La mayor desventaja es el tiempo y esfuerzo que consume el hecho de desarrollar y validar una biblioteca NIR, a diferencia del Raman, que este proceso es mucho más sencillo y lo puede llevar a cabo personal no experto. Al mismo tiempo, esto se puede convertir en una desventaja si empiezan a aparecer ambigüedades entre métodos, ya que el software es cerrado. Así ambas técnicas presentan sus ventajas y desventajas, y debido a sus características, son complementarias.

4.5. BIBLIOGRAFÍA

1. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Capítulo 5 Producción. 11 (2015).
2. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario Anexo 8: Toma de muestras de materiales de partida y de acondicionamiento. 3 (2010).
3. Bogdanich, W. & Hooker, J. From China to Panama, a Trail of Poisoned Medicine. *New York Times* (2007). Available at: <https://www.nytimes.com/2007/05/06/world/americas/06poison.html>.
4. Rentz, D. E. *et al.* Outbreak of acute renal failure in Panama in 2006: a case-control study. *World Health Organization (WHO)* (2008).
5. Miró Vera, A. Y. & Alcalà Bernàrdez, M. Near-Infrared Spectroscopy in Identification of Pharmaceutical Raw Materials. *Encycl. Anal. Chem.* 1–19 (2017).
6. Rowe, R. C. R., Sheskey, P. J. S. & Cook, W. Handbook Pharmaceutical Excipients, Sixth Edition. 1064 (2009). doi:<http://dx.doi.org/10.1016/B978-012088479-7/50039-0>
7. McCreery, R. L., Horn, A. J., Spencer, J. & Jefferson, E. Noninvasive identification of materials inside USP vials with Raman spectroscopy and a Raman spectral library. *J. Pharm. Sci.* **87**, 1–8 (1998).
8. Blanco, M. & Romero, M. A. Near-infrared libraries in the pharmaceutical industry: A solution for identity confirmation. *Analyst* **126**, 2212–2217 (2001).
9. Sun, L. *et al.* Pharmaceutical Raw Material Identification Using Miniature Near-Infrared (MicroNIR) Spectroscopy and Supervised Pattern Recognition Using Support Vector Machine. *Appl. Spectrosc.* **70**, 816–825 (2016).
10. Casian, T. *et al.* Development, validation and comparison of near infrared and Raman spectroscopic methods for fast characterization of tablets with amlodipine and valsartan. *Talanta* **167**, 333–343 (2017).
11. Esmonde-White, K. A., Cuellar, M., Uerpmann, C., Lenain, B. & Lewis, I. R. Raman spectroscopy as a process analytical technology for pharmaceutical manufacturing and bioprocessing. *Anal. Bioanal. Chem.* **409**, 637–649 (2017).
12. Reich, G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug Deliv. Rev.* **57**, 1109–1143 (2005).
13. Jestel, N. L. Chapter 7 - Raman Spectroscopy. in *Process Analytical Technology* 195–243 (2010).
14. Chalmers, J. M., Edwards, H. G. M. & Hargreaves, M. D. *Infrared and Raman Spectroscopy in Forensic Science.* (2012).
15. Rodionova, O. Y., Sokovikov, Y. V & Pomerantsev, A. L. Quality control of packed raw materials in

- pharmaceutical industry. *Anal. Chim. Acta* **642**, 222–227 (2009).
16. E. Bertoft, A. Analysing starch structure. in *Starch in Food: Structure, Function and Applications* 57–96 (2004). doi:10.1016/B978-1-85573-731-0.50006-2
 17. Thoorens, G., Krier, F., Leclercq, B., Carlin, B. & Evrard, B. Microcrystalline cellulose, a direct compression binder in a quality by design environment - A review. *Int. J. Pharm.* **473**, 64–72 (2014).
 18. Lerk, C. F. Consolidation and compactation of lactose. *Drug Dev. Ind. Pharm.* **19**, 2359–2398 (1993).
 19. Dégardin, K., Guillemain, A. & Roggo, Y. Comprehensive Study of a Handheld Raman Spectrometer for the Analysis of Counterfeits of Solid-Dosage Form Medicines. *J. Spectrosc.* **2017**, (2017).
 20. United States Pharmacopeia. USP 40–NF 35 <1120> Raman spectroscopy.

5. IDENTIFICACIÓN DE PRODUCTOS GRANEL

5.1. INTRODUCCIÓN

El producto granel es aquel que ya ha pasado por todas las fases de producción excepto el acondicionamiento final¹. Cuando este producto es adquirido como tal por parte de la compañía farmacéutica, debe tratarse como un material de partida², por lo que es necesario verificar la identidad del mismo.

Como en el caso de las materias primas, la técnica analítica NIR, en combinación con un método de reconocimiento de pautas, puede ser útil para este propósito, y así evitar tener que emplear técnicas que consuman tiempo y requieran una preparación previa de la muestra. La elección del PRM se hará en base a la disponibilidad de algoritmos que ofrezca el software quimiométrico del instrumento, así como de la dificultad de la señal que presente la muestra.

En muchas ocasiones, un mismo medicamento se formula en diversas dosis. Por lo que no solo es necesario identificar el producto, sino que se debe cualificar. Esto es, el método analítico tiene que ser capaz de diferenciar las dosis. En comprimidos o cápsulas, las diferentes dosis se consiguen modificando la concentración de la sustancia activa o bien variando la cantidad unitaria de una misma formulación (fórmulas proporcionales). En el primer supuesto, el modelo quimiométrico que se desarrolle estará basado en las diferencias de concentración de API entre dosis. En el segundo, dado que la concentración es la misma, es necesario desarrollar un modelo que además de contener información sobre la identidad química, sea capaz de diferenciar por características físicas. El NIR es una técnica útil para ello, ya que permite diferenciar comprimidos, por ejemplo, según su dureza o porosidad³⁻⁶. Esta capacidad es debida a que la propagación de la luz en la muestra se ve afectada por las condiciones de empaquetamiento de la misma³. Los trabajos referenciados han evaluado la capacidad de los modos de adquisición en reflectancia difusa y transmitancia para este fin. Otro factor que tiene una gran influencia en los espectros adquiridos en transmisión es el grosor tanto del comprimido como del recubrimiento^{7,8}.

En el laboratorio de control de calidad se utilizaba un método de identificación y cualificación por NIR de tres presentaciones de comprimidos con una misma sustancia activa, alternativo a los métodos HPLC definidos en los correspondientes expedientes autorizados. El modelo NIR carecía de robustez y frecuentemente no identificaba el producto o lo cualificaba como una presentación incorrecta. El objetivo de este trabajo ha sido encontrar la causa raíz de los errores de identificación que presentaba el método y proponer una alternativa viable para la correcta identificación de los mismos.

5.2. MÉTODO EXPERIMENTAL

5.2.1. Muestras

Los productos sujetos de estudio son comprimidos recubiertos que tiene como principio activo el Olmesartan Medoxomilo. Se comercializan medicamentos en tres presentaciones de 10, 20 y 40 mg. En los tres comprimidos la sustancia activa se encuentra en la misma concentración (9.5%), consiguiéndose cada dosis de 10, 20 y 40 mg con una cantidad unitaria distinta.

El número de lotes de cada presentación que se usaron para la calibración/optimización y la validación externa del método de identificación se detallan en la Tabla 5.1. Los espectros de calibración fueron escogidos basándose en el análisis visual de los mismos una vez aplicado el SNV como pretratamiento y escogiendo los de mayor variabilidad (mayor, menor y absorción media) para cada presentación; todos los demás se destinaron a la validación. Además, para la validación externa también se usaron núcleos de cada una de las dosis y otros comprimidos suministrados por el mismo proveedor del producto granel, que contienen, además del Olmesartan Medoxomilo, otros dos APIs.

Tabla 5.1 Número de lotes usados para la calibración y validación del método de identificación de comprimidos de Olmesartan.

Producto	Calibración	Validación
Olmesartan 10mg	5 lotes	30 lotes
Olmesartan 20mg	5 lotes	91 lotes
Olmesartan 40mg	5 lotes	63 lotes

Cada lote consta de diversos envases, por lo que se seleccionaron diez comprimidos de la muestra representativa para registrar el espectro NIR. Todos los lotes fueron identificados por el método de referencia.

5.2.2. Adquisición de espectros

La adquisición de los espectros se llevó a cabo con el espectrómetro FT-NIR Nicolet Antaris II (Thermo Scientific, USA) equipado con detector InGaAs. Se usó la esfera integradora para la adquisición en reflectancia difusa y el módulo estándar de transmisión para comprimidos para la adquisición en transmisión. El instrumento se controla con el software Result (Thermo Scientific, USA).

Para obtener los espectros en reflectancia difusa, los comprimidos se colocaban en la esfera integradora centrándolos con un iris. Una vez abierto el iris, se registraban 32 *scans* en el rango espectral 10000-4000 cm^{-1} con una resolución de 16 cm^{-1} . La referencia se obtenía de una pieza de oro interna.

Para obtener los espectros en transmisión, los comprimidos se colocaban en la esfera integradora con la ayuda de un carrusel y se cubrían con el módulo. Cada espectro adquirido fue el resultado de 100 *scans* en el rango espectral 10000-6000 cm^{-1} y con una resolución de 32cm^{-1} . La referencia se adquiría sin ninguna muestra en el módulo.

5.2.3. Tratamiento de datos

Para la exploración de los datos, tanto históricos como actuales, se empleó el software The Unscrambler X (Camo Analytics, Noruega). Distintos pretratamientos y rangos fueron seleccionados para aplicar un análisis de componentes principales (PCA) que permitiera detectar clústeres o tendencias.

Para la calibración y validación del modelo se usó el software TQ Analyst (Thermo Scientific, USA). Las clases incluidas en la biblioteca fueron las tres presentaciones de Olmesartan 10, 20 y 40 mg. Sin ningún pretratamiento de los datos e incluyendo el rango espectral $9800-6300\text{cm}^{-1}$, se escogió como algoritmo de reconocimiento de pautas el análisis discriminante. Este se basa en reducir las variables mediante y calcular la distancia Mahalanobis entre los *scores* y el centro de la clase. El umbral establecido para cada una de las clases fue una distancia de 2.3, dado que era la que permitía identificar correctamente todos los productos evitando que hubieran confusiones – falsos positivos- entre clases. Una nueva muestra se clasificaría al grupo con el cual presente una distancia menor al umbral.

5.3. RESULTADOS Y DISCUSIÓN

5.3.1. Cualificación por modo reflectancia

Los comprimidos que se deben identificar y cualificar son de tres dosis distintas, pero para fabricar cada una de ellas, se utiliza el mismo granulado. Así, la concentración de API es la misma para todas, diferenciándolas solo la cantidad dosificada y por ende, las propiedades físicas. La diferencia de la dureza, la forma y el grosor es lo que permitiría cualificarlas. En la Tabla 5.2 se resumen las características y formas de cada una de las presentaciones.

Tabla 5.2 Característica de las tres presentaciones disponibles de comprimidos de Olmesartan.

Características	Olmesartan 10mg	Olmesartan 20mg	Olmesartan 40mg
Forma			
Dimensiones (mm)	6.5	8.5	15.0x7.0
Grosor (mm)	3.2	3.8	4.8
Dureza (kp)	9	12	16

El espectro del API adquirido en reflectancia difusa presenta bandas en todo el rango espectral, siendo las más intensas en la zona de 4000 a 5000 cm^{-1} . Cuando se compara con el espectro de un comprimido de 20mg (Figura 5.1), a priori, es difícil detectar la contribución del API. Al aplicar la primera derivada se observa que, a parte de las bandas más intensas mencionadas anteriormente, también hay una contribución del resto de bandas que aparecen entre 6000 y 9000 cm^{-1} .

El método NIR desarrollado hasta la fecha se basaba en el registro de espectros por reflectancia difusa de los comprimidos dispuestos en el iris de la esfera integradora, sin ningún elemento que los cubriese. El modelo cualitativo se basaba en el análisis discriminante con el cálculo de distancia Mahalanobis de los espectros sin aplicar ningún pretratamiento. El primer paso fue determinar el motivo de la falta de selectividad y robustez del modelo. Dado que los espectros eran adquiridos en modo de reflectancia, cabe esperar que la diferencia entre cada clase sea debido al fenómeno de *scattering*.

Los espectros incluidos en la calibración del modelo se representan en la Figura 5.2. Para cada presentación se observa un desplazamiento de la línea base, siendo mayor la diferencia entre la presentación de 20mg de las demás. Cuando se realizó un análisis por componentes principales (PCA) de los mismos espectros se obtuvo el gráfico de *scores* de la Figura 5.2. En él se observa que el primer componente principal (PC1) separaba las muestras según la presentación de comprimidos, confirmándose lo que se detectaba en los espectros. Pero aunque se define un clúster para cada clase, la separación entre ellos es muy ambigua, llegando a tener muestras solapadas entre ellos. Esto provoca que la selectividad se vea comprometida.

El modo de adquisición de espectros en este método, como se ha mencionado anteriormente, era el de reflectancia. La muestra se irradia con un haz de luz y parte de esta luz es reflejada. Dicha reflexión puede ser especular, que no aporta información de la muestra, o bien difusa. Esta última tiene lugar en todas las direcciones y es debido a los procesos absorción y dispersión. Esta es la que nos aporta la información de la muestra. En la Figura 5.3 se presenta un esquema de la lectura de comprimidos. Esta adquisición se llevaba a cabo sin cubrir los comprimidos.



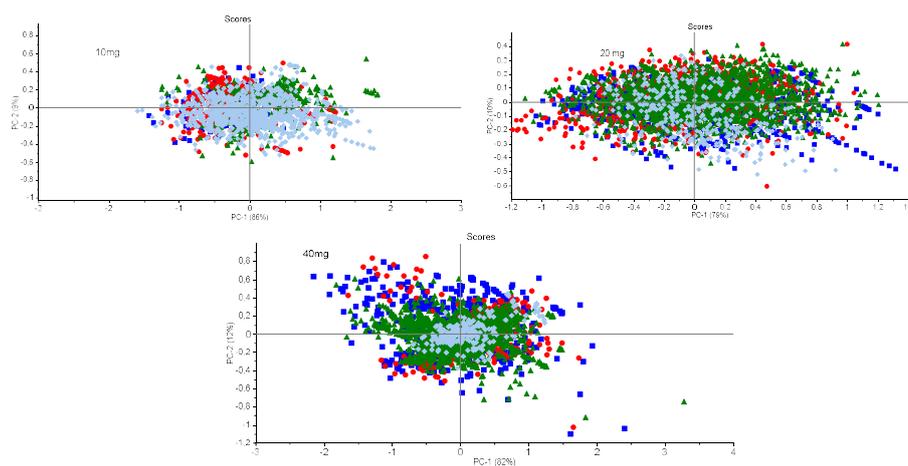
Al tomarse los espectros al aire, sin ser protegidos de la luz ambiental, esta tiene una influencia sobre el espectro que se obtiene. Se debería tapar con un material negro para evitar la entrada de luz exterior y para que la radiación “perdida” del haz incidente -la que no interactúa con la muestra- fuera absorbida por ese material, y no influenciara en el espectro de la muestra.

Cuando la adquisición de los espectros se realizó por reflectancia y cubriendo los comprimidos con una cerámica negra, se obtuvieron los espectros de la Figura 5.4. En este caso, únicamente hay información referente a la muestra y se puede observar cómo los espectros de los comprimidos correspondientes a las presentaciones de 10 y 40 mg se solapan completamente. El gráfico de *scores* de la Figura 5.4 confirma dicho solapamiento.

Esto demuestra que la diferencia entre clases que se observaba en la Figura 5.2 no era solamente debido a la diferencia entre dichas clases, sino que también se veía influenciada por el entorno. Esto es lo que provocaba la falta de selectividad del método. Se puede concluir que, en este caso, el

modo de reflectancia no es un método adecuado para cualificar los comprimidos de manera inequívoca entre las distintas presentaciones.

Otro parámetro que se tuvo en consideración fue la reproducibilidad de la medida debido a las lámparas. Dado que el modelo se desarrolló sin ningún pretratamiento de los espectros, era necesario evaluar la influencia que podía tener el cambio de lámpara del instrumento NIR. Para ello se recopilaron los espectros de los comprimidos analizados durante cuatro años. Como la lámpara, por procedimiento, se cambia anualmente, esto equivale a tener información obtenida con cuatro lámparas distintas. Los gráficos de *scores* obtenidos para cada presentación (Figura 5.5) no presentaron ninguna tendencia temporal, por lo que se concluyó que las lámparas daban señales estables.



5.3.2. Cualificación por modo transmisión

Era necesario proponer un método alternativo que fuera capaz de identificar y cualificar los comprimidos de Olmesartan de manera inequívoca. Para ello se evaluó el modo de adquisición de espectros mediante transmisión. En la Figura 5.6 se muestran los espectros de los comprimidos de Olmesartan. Entre las tres presentaciones se observan diferencias en los espectros, por lo que cabe esperar que se puedan clasificar en distintos grupos utilizando un análisis multivariable. Las diferencias que se observan en los espectros obtenidos en cada presentación seguramente se deban a un conjunto de efectos. La diferencia en la línea base puede ser debida a los distintos procesos de *scattering* producidos por la diferencia en la dureza de los comprimidos y/o a la distinta porosidad de los mismos. Por otro lado, la diferencia en los contornos o en la absorción relativa de las bandas puede ser consecuencia de los distintos grosores de los comprimidos. Al aumentar el grosor, aumenta el camino óptico recorrido por la radiación, por lo que la muestra absorbe más luz y llega menos al detector. Esto se traduce en una disminución de la información presente en el espectro.

El gráfico de *scores* obtenido con el PCA mostrado en la Figura 5.7, pone de manifiesto que el primer PC, que explica el 99.6% de la varianza, sirve para separar los comprimidos de 10mg del

resto y que es el segundo PC, con un 0.4% de varianza explicada, que permite diferenciar los de 20 y 40 mg.

Con los espectros de calibración se desarrolló un modelo utilizando el algoritmo de análisis discriminante. El modelo se calculó sin aplicar ningún pretratamiento a los espectro NIR e incluyendo el rango espectral 9800-6300 cm^{-1} . Los resultados de la validación interna permitieron fijar una distancia de 2.3 como umbral común para todas las clases. Los gráficos de las distancias de Mahalanobis obtenidos para cada pareja de clases (Figura 5.8) confirmaban la habilidad del modelo y del valor de aceptación de discriminar entre las muestras según su dosificación. Todas las muestras de calibración se clasificaron correctamente sin presentar ninguna ambigüedad.

La validación externa se llevó a cabo con todos los lotes que no se usaron en la calibración del modelo. En la Figura 5.9 se resumen los resultados que se obtuvieron en todos los lotes analizados de Olmesartan. Cabe destacar que solo se presenta el resultado de un comprimido de los diez analizados por lote para simplificar los gráficos. Con ellos se demuestra la robustez del modelo y la capacidad de clasificar correctamente los productos en el grupo que pertenecen.

La selectividad del modelo se demostró no solo con la correcta identificación del producto, sino que también con núcleos y otro medicamento que contiene, junto al Olmesartan Medoxomilo, otros dos APIs. Los resultados obtenidos y presentados en la Tabla 5.3 corroboran la capacidad de rechazar productos que no pertenecen al producto objetivo.

Tabla 5.3 Resultados de la validación externa realizada con núcleos y 5 productos. El producto A, B, C, D y E contiene los mismos 3 APIs en distintas proporciones.

Producto	Clase más cercana	Distancia	Resultado
Núcleo 10mg	10mg	4.66	Fallo
Núcleo 20mg	20mg	3.73	Fallo
Núcleo 40mg	20mg	5.00	Fallo
Producto A	20mg	4.64	Fallo
Producto B	20mg	10.98	Fallo
Producto C	40mg	7.94	Fallo
Producto D	40mg	10.63	Fallo
Producto E	40mg	8.68	Fallo

5.4. CONCLUSIONES

Se ha determinado que la reflectancia difusa no es el método adecuado para cualificar el producto objetivo según la dosis de los comprimidos objetos de este estudio. Una vez eliminado el factor de

la luz ambiental, se ha demostrado la completa falta de selectividad del mismo. En consecuencia, se ha propuesto un método alternativo registrando los espectros en transmisión, traduciéndose en ellos las diferencias físicas que presenta cada una de las presentaciones. El modelo obtenido presenta una buena selectividad frente a los núcleos del mismo producto y a comprimidos de otros medicamentos, y también una buena robustez, habiendo identificado y cualificado correctamente un gran número de lotes.

Ahora bien, no se puede perder de vista que en el caso de la presentación con mayor grosor, la información del espectro es bastante pobre, debido a que la radiación que llega al detector es baja. En consecuencia, se propone una mejora en el método. Esta consiste en adquirir automáticamente dos espectros por cada comprimido analizado, en reflectancia y transmisión. El primero serviría para identificar, en una primera fase, el comprimido como Olmesartan. La cualificación del mismo, esto es, determinar a la dosis o presentación a la que pertenece, se realizaría en una segunda fase con el espectro de transmisión

5.5. BIBLIOGRAFÍA

1. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Glosario. (2010).
2. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Capítulo 5 Producción. 11 (2015).
3. Ito, M. *et al.* Development of a method for the determination of caffeine anhydrate in various designed intact tablets by near-infrared spectroscopy: A comparison between reflectance and transmittance technique. *J. Pharm. Biomed. Anal.* **47**, 819–827 (2008).
4. Burns, D. A. & Ciurczak, E. W. *Handbook of Near-Infrared Analysis*. (CRC Press Taylor and Francis, 2008).
5. Casian, T. *et al.* Development, validation and comparison of near infrared and Raman spectroscopic methods for fast characterization of tablets with amlodipine and valsartan. *Talanta* **167**, 333–343 (2017).
6. Blanco, M., Alcalá, M., González, J. M. & Torras, E. A Process Analytical Technology Approach Based on Near Infrared Spectroscopy: Tablet Hardness, Content Uniformity, and Dissolution Test Measurements of Intact Tablets. *J. Pharm. Sci.* **95**, 2137–2144 (2006).
7. Xiang, D. *et al.* Evaluation of transmission and reflection modalities for measuring content uniformity of pharmaceutical tablets with near-infrared spectroscopy. *Appl. Spectrosc.* **63**, 33–47 (2009).
8. Ulmschneider, M. & Pénigault, E. Non-invasive confirmation of the identity of tablets by near-infrared spectroscopy. *Analisis* **28**, 336–346 (2000).

Parte III

ANÁLISIS EN EL PROCESO

6. IMPLEMENTACIÓN DE UN SISTEMA PAT EN LA INDUSTRIA FARMACÉUTICA

6.1. INTRODUCCIÓN

Cuando la espectroscopia NIR se combina con un sistema de control en proceso puede emplearse como una herramienta PAT para analizar, de forma más o menos automatizada, los CQAs de un producto o bien, para seguir un proceso y determinar el punto final de una operación unitaria. Esto permite trasladar la función de un laboratorio de control de calidad a la planta de producción. Para implementar satisfactoriamente un sistema PAT es necesario elaborar un proyecto y que toda la organización esté implicada en él. En la bibliografía se encuentran distintas propuestas de las fases que deberían formar parte de este tipo de proyectos. Aunque cada autor utiliza nombres diferentes, en el fondo, el objetivo y las tareas a realizar son las mismas. En la Figura 6.1 se presenta un ejemplo.

En la Tabla 6.1 se resumen los objetivos y algunas de las tareas principales a realizar en cada una de las etapas mencionadas cuando la finalidad es implementar el NIR como herramienta PAT. Es necesario tener presente que se está trabajando en un entorno regulado, por lo que se debe cumplir con lo establecido en las GMPs.

Tabla 6.1 Etapas a seguir para el desarrollo de un proyecto PAT usando como analizador un instrumento NIR.

Etapas	Objetivo y/o tareas¹
Definición del proyecto	- Definir el objetivo del proyecto y realizar todas las tareas propias de puesta en marcha de un proyecto (cronogramas, asignar los recursos destinados...).
Alcance del proyecto	- Recopilar detalles del proceso, definir requisitos analíticos y desarrollar un conocimiento del proceso (operaciones unitarias, tendencias históricas, identificar fuentes de variabilidad conocidas y potenciales...) en base al

Tabla 6.1 Etapas a seguir para el desarrollo de un proyecto PAT usando como analizador un instrumento NIR.

Etapa	Objetivo y/o tareas ¹
	cual se desarrollará el sistema PAT.
Prueba de concepto	<ul style="list-style-type: none"> - Buscar soluciones PAT (en la bibliografía y en el mercado) que den información en tiempo real, relacionada con las necesidades definidas en el objetivo. Decidir qué tecnología usar, según el conocimiento interno o la madurez de la misma. - Realizar las pruebas necesarias para evaluar la capacidad del instrumento a dar la respuesta de interés en el rango especificado. Elegir un instrumento. - Estudio de viabilidad con el instrumento en las instalaciones y con las modificaciones necesarias en los equipos. En esta etapa se optimiza la medición y se desarrollan los modelos quimiométricos.
Diseñar el sistema	<ul style="list-style-type: none"> - Diseñar el sistema de análisis. Esto debe incluir, pero no está limitado, la siguiente información: <ul style="list-style-type: none"> o Detallar cómo se realizará el muestreo o Detallar cómo se llevará a cabo el análisis o Detallar cómo se gestionarán los datos - Estimar los costes. - Adquirir el analizador (NIR) si es necesario. Realizar las modificaciones pertinentes a los equipos de proceso.
Implementación industrial	<ul style="list-style-type: none"> - Instalar el analizador (NIR). Realizar la cualificación de instalación y operación (IQ, OQ). - IQ y OQ del software quimiométrico. - Comprobación del buen funcionamiento de los modelos quimiométricos. Si no es así, rehacer los modelos con la variabilidad del proceso. - Establecer los parámetros de control para la comprobación del buen funcionamiento del modelo quimiométrico. - Implementar el sistema de automatización y gestión de los datos. Cualificar dicho sistema. - Redactar las instrucciones de trabajo y toda la documentación relacionada. Formación del personal sobre el sistema de análisis. - Realizar la cualificación del funcionamiento de todo el sistema (PQ). - Una vez demostrado y documentado el buen funcionamiento del sistema PAT se puede implementar en rutina. - Registro y espera de la aprobación, si es necesario.
Implementación en rutina	<ul style="list-style-type: none"> - Verificar de manera periódica la validez del modelo quimiométrico. La frecuencia se determina según la robustez demostrada en la validación del mismo. Se suele aceptar 10 lotes. - Si hay algún cambio que pueda afectar en la señal NIR, ya sea por el material de partida o los parámetros de producción, se debe realizar un control de cambios para que el departamento encargado de mantener el modelo pueda realizar las acciones necesarias. - Realizar el mantenimiento necesario en el instrumento.
Gestión del ciclo de vida / Mejora continua	<ul style="list-style-type: none"> - Evaluar periódicamente la necesidad de mantener el sistema PAT tal como se concibió en un inicio. Realizar las mejoras necesarias acorde con la evolución en la fabricación o en la tecnología.

El objetivo de este trabajo ha consistido en implementar industrialmente un método de análisis mediante NIR para determinar en tiempo real los CQAs de un producto granulado, una vez acabada

la etapa de mezclado. Esto permitiría la inmediata liberación interna del granel para poder pasar a la siguiente etapa de dosificación. Las primeras fases del proyecto ya se habían llevado a cabo por parte de Juan G. Rosas et al.² y Laboratorios Menarini. Desde un inicio, se detectó que los modelos realizados hasta la fecha no presentaban la robustez suficiente para poderlos usar en rutina. Por esto, el objetivo también ha sido desarrollar modelos quimiométricos capaces de predecir de forma robusta los CQAs (identificación, cuantificación y contenido en agua) en el entorno de producción. Antes de proceder a exponer el trabajo realizado, es imprescindible poner en contexto el proyecto.

6.2. ANTECEDENTES

El objetivo de este proyecto era implementar un sistema PAT que permitiera determinar, en tiempo real, los CQAs necesarios para liberar un producto una vez finalizada la etapa de mezcla; para aumentar la productividad y optimizar recursos. El medicamento es un producto farmacéutico granulado que se comercializa en sobres. El proceso de producción se esquematiza en la Figura 6.2. Una vez terminada la granulación de una parte de la formulación mediante lecho fluido, se le añade un último excipiente extra-granular. Es necesario realizar una mezcla para poder homogeneizarlo. En este caso, tanto el producto granulado como el último excipiente, se descargan en un depósito denominado comúnmente como bin. El proceso de mezclado tiene como parámetros críticos de proceso la velocidad (rpm) y el tiempo (min) de mezcla. Estos parámetros vienen fijados por una validación de dicho proceso. La forma tradicional de proceder, cuando la mezcla ha finalizado, es transportar este bin a una sala controlada para realizar un muestreo por niveles del producto. Esta muestra es la que se analiza en el departamento de control de calidad para determinar los CQAs que permiten liberar el producto y que éste pase a la siguiente etapa de producción, la dosificación.

La implementación de un sistema de medida en el sitio de mezcla llevaría a la liberación en tiempo real. En el año 2012, por parte de Laboratorios Menarini, se adquirió un NIR capaz de registrar espectros de forma no invasiva. Desde entonces, se ha instalado un software que permite la

monitorización en tiempo real, es decir adquirir los espectros, predecir los CQA y enviar los valores. Una de las bocas estándar de los depósitos bin se han modificado incorporando una ventana de zafiro (de 3mm de grosor y 2cm de diámetro) y un adaptador que permite el anclaje de la sonda NIR. Además, se ha ideado una aplicación que permite la gestión de los datos que se generaran en las predicciones. Por parte del Grupo de Quimiometría Aplicada de la UAB se desarrollaron modelos quimiométricos capaces, entre otros, de identificar el granulado, así como de predecir el riqueza del API y el contenido en agua². El esquema del sistema instalado se encuentra en la Figura 6.3.

El sistema funcionaba de la siguiente manera. Una vez finalizada la mezcla, el bin quedaba posicionado con la boca modificada abajo, permitiendo que el producto caiga sobre la misma. Desde el PC se daba la orden para que empezara la adquisición de los espectros con el bin estático. Automáticamente, desde el software de monitorización en tiempo real se predecían los valores correspondientes a los espectros adquiridos según los modelos introducidos previamente. En 2016, se había finalizado la fase de diseño del sistema y se había realizado alguna tarea de industrialización, pero era necesario realizar prácticamente todas las acciones de implementación.

6.3. MÉTODO EXPERIMENTAL

6.3.1. Producto

El producto es un granulado que contiene Nimesulide como API, sacarosa como excipiente mayoritario (aproximadamente un 90% en peso) y ácido cítrico, esencia de naranja, maltodextrina

y macrogol como excipientes minoritarios. La concentración de API en el producto final es de 50mg/g (5% en peso). Las especificaciones de los CQAs que se deben determinar con el método NIR son las que se demuestran en la Tabla 6.2.

Tabla 6.2 Especificaciones del producto granulado.

CQA	Especificación
Identificación	Pasa
Contenido de API (mg/g)	47.5-52.5
Coefficiente de variación del API (%)	≤5%
Contenido en agua (% m/m)	≤0.5%

6.3.2. Conjunto de muestras para la calibración y validación

6.3.2.1. Modelo de identificación

Las muestras usadas para la calibración y validación del método de identificación del producto Nimesulide fueron las propias muestras de producción, los excipientes de la formulación, el placebo, y un producto que se almacena en los mismos bins una vez se ha granulado. Dado que los espectros NIR son sensibles a las propiedades físicas del producto, el placebo se granuló. En la Tabla 6.3 se detallan los usos y los motivos que llevaron a incluir cada uno de las muestras en el estudio.

Tabla 6.3 Muestras incluidas en la calibración y validación del modelo de identificación y la justificación de ser incluidas.

Muestras incluidas	Justificación
Producto granulado Nimesulide	Calibración, optimización y validación externa (especificidad y robustez). Es el producto objetivo de la identificación.
Sacarosa	Validación externa, para demostrar la especificidad. Excipiente mayoritario
Ácido cítrico	
Esencia de naranja	Validación externa, para demostrar la especificidad. Excipiente minoritarios
Maltodextrina	
Macrogol	
Placebo (polvo y granulado)	Validación externa, para demostrar la especificidad. Producto objeto de identificación sin presencia de API.
Otro producto granulado	Validación externa, para demostrar la especificidad. Producto que se manipula cerca del Nimesulide en la planta de fabricación.

6.3.2.2. Modelo para la cuantificación de API

Para el desarrollo de un modelo cuantitativo capaz de predecir la riqueza de API se usaron dos tipos de muestras, las de producción y las preparadas en el laboratorio. Las primeras servirían para introducir la variabilidad esperada del proceso. Las segundas servirían para expandir el rango de

concentración hasta un $\pm 30\%$ respecto el valor nominal (35-65 mg/g). La estrategia usada para expandir el rango de concentración fue el de dopar muestras de producción, con API o placebo, según si se quería aumentar o disminuir la concentración de API.

En total se prepararon 57 muestras de laboratorio, que englobaban 7 niveles de concentración (70, 80, 95, 105, 115 y 130 % respecto al valor nominal). Primero se prepararon 3 placebos con distintas proporciones de cada uno de los excipientes (variando las concentraciones un $\pm 5\%$ respecto al valor nominal) para disminuir las correlación entre los componentes de la formulación en las muestras finales. Después se seleccionaron 6 lotes del producto granulado de Nimesulide de producción, que servirían como base para preparar las muestras de laboratorio. A partir del producto base, se añadió cierta cantidad de API o placebo (en la Tabla 6.4 se detallan las cantidades de cada uno de las sustancias usadas para preparar 100g de cada nivel de concentración). El placebo usado para la preparación de cada muestra se adjudicó de manera aleatoria. El producto final se mezcló en un bombo V de la planta piloto. 45 muestras se emplearon para la calibración y optimización del modelo y 12 para la validación externa del mismo. Los lotes en ambos conjuntos eran independientes.

Tabla 6.4 Cantidades usadas de producto Nimesulide de producción, API y placebo usadas para preparar cada uno de los niveles de concentración.

Nivel de concentración nominal de API (%)	Concentración de API (mg/g en peso)	Cantidad de muestra de producción (g)	Cantidad de placebo (g)	Cantidad de API (g)
70	35.0	70.00	0.00	30.00
85	42.5	85.00	0.00	15.00
95	47.5	95.00	0.00	5.00
100	50.0	100.00	0.00	0.00
105	52.5	99.74	0.26	0.00
115	57.5	99.21	0.79	0.00
130	65.0	98.42	1.58	0.00

6.3.2.3. Modelo cuantitativo para la determinación del contenido en agua

Como en el caso anterior, se usaron muestras de producción y muestras preparadas en el laboratorio. Las muestras de producción se seleccionaron de manera que incluyeran toda la variabilidad esperada procedente del proceso. Para aumentar el rango de humedad, se prepararon 47 muestras en el laboratorio. Dado que la especificación interna del producto establece que el contenido en agua tiene que ser menor o igual al 0.50%, y el valor obtenido en producción suele ser 0.20-0.30 %, se estableció un protocolo que permitiera expandir el rango de 0.10 a 1.50 %. Para ello se usaron los mismos 6 lotes de producción seleccionados para la preparación de muestras para expandir la concentración de API. Para secar el producto, las muestras se pusieron en una estufa a 50°C, dejando un tiempo de 2.5h para conseguir un porcentaje cercano al 0.10%. Cabe mencionar

que se realizó un estudio previo con el que se concluyó que no era posible disminuir el contenido de agua por debajo de 0.1%. Para humedecer el producto, se puso en la estufa un recipiente con agua a 40°C, de manera que se creara un ambiente húmedo. El producto se introdujo en el horno durante tiempos variables para lograr un aumento en el contenido de agua. En la Tabla 6.5 se detallan los tiempos y los contenidos de agua esperados. La cantidad preparada de cada muestra fueron 100g. 37 muestras se incluyeron en el conjunto de calibración y optimización, y las 10 restantes en el conjunto de validación externa. Los lotes de cada conjunto eran independientes.

Tabla 6.5 Tiempo que se expuso el producto Nimesulide de producción a un ambiente húmedo y el contenido en agua esperado.

Contenido en agua (%)	Tiempo expuesto a humedad (h)
0.40	0.5
0.80	2.0
1.20	4.1
1.50	4.4

6.3.3. Adquisición de espectros

El equipo que se usó para la adquisición de los espectros fue un LabSpec Pro 2500 (ASD inc., USA) con un rango espectral de 350 a 2500 nm. El principio de dispersión en la zona del visible-NIR (350-1000 nm) es una matriz de diodos con filtros y en la región del NIR cuenta con dos monocromadores con rejilla holográfica cóncava (uno entre 1000 y 1800 nm y el otro entre 1800 y 2500 nm). Este instrumento también cuenta con tres detectores. En la zona del visible-NIR tiene una matriz de fotodiodos con 512 canales. En la zona del NIR tiene dos detectores InGaAs enfriados termo-eléctricamente. La resolución de los espectros es variable en todo el rango espectral, siendo de 4 y 10 nm en la región del visible y del NIR, respectivamente. El número de *scans* para la adquisición de un espectro fueron 32. El accesorio de muestreo para registrar los espectros en reflectancia difusa, fue una sonda modificada que permitía el anclaje con el tapón del bin y que estaba unida al equipo mediante una fibra óptica de 3m. Antes de cada medida se registraba la referencia con una placa de *spectralon*.

El equipo fue cualificado según el Anexo 15 de las GMPs³. Antes de proceder con las lecturas, el equipo se verifica diariamente según la farmacopea europea⁴. El test consiste en verificar la exactitud de la longitud de onda, la linealidad fotométrica y el nivel de ruido espectrofotométrico. Para llevarlo a cabo se utiliza un kit de 7 estándares externos (Avian Technologies, USA) que presentaban distintos porcentajes de reflectancia.

El registro de los espectros de las muestras de producción y de las muestras preparadas en el laboratorio fue distinto, debido a la cantidad de muestra disponible. Para las primeras, había 400Kg

de producto, mientras que para las segundas, la cantidad era de 100g. A continuación se especifica el método para cada una de ellas.

- Muestras de producción: una vez finalizado el proceso de mezcla, el bin quedaba posicionado hacia abajo, para que el producto granulado cubriera por completo la ventana de zafiro. La referencia se registró con una placa de *spectralon*. Acto seguido, la sonda se acopló al tapón del bin con la ventana de zafiro, y se registró el espectro. Una vez finalizado, se extraía la sonda y se procedía a dar una vuelta al bin, lo que permitía renovar la muestra analizada. Para cada lote, se repetía este proceso por triplicado, obteniendo tres espectros.
- Muestras de laboratorio: en este caso, las muestras no se podían analizar en el bin, debido a la cantidad disponible. La metodología que se siguió fue la de poner los 100g preparados de cada muestra en un tapón de bin, asegurando que, aparte de cubrir la ventana de zafiro, no se perdía radiación durante las medidas. El registro del espectro de referencia y de la muestra se realizó de la misma manera que en el caso de las de producción. Entre espectro y espectro, se mezclaba manualmente el producto de manera que cambiase la muestra analizada. En total se obtuvieron 3 espectros por cada muestra de laboratorio.

6.3.4. Métodos de referencia

Dado que el NIR es una técnica indirecta, los valores de referencia se obtuvieron con los métodos analíticos validados por Laboratorios Menarini y autorizados para el control inicial del medicamento.

La identificación y el contenido de API se obtuvieron mediante un método de cromatografía líquida de alta presión (HPLC) usando un cromatógrafo serie 1200 con un detector de red de diodos (Agilent Technology, USA) equipado con un columna C18 de dimensiones 4.6x50mm. La fase móvil usada fue agua acidificada y metanol (40/60 V/V). Los cromatogramas se procesaron con el software Empower (Waters, USA). El análisis se realizó por triplicado, por lo que el promedio se consideró la concentración de API del lote (o de la muestra de laboratorio), mientras que el coeficiente de variación se relacionó con la homogeneidad del lote.

Los valores de referencia para la determinación del contenido en agua se obtuvieron por el método de pérdida por desecación utilizando un analizador de humedad (Mettler Toledo, USA) a 90°C durante 5 minutos. En este caso, solo se realizó un análisis de la muestra representativa del lote (o de la muestra de laboratorio).

6.3.5. Software

Los softwares implicados en este trabajo han sido tres. Para la adquisición de los espectros de las muestras de laboratorio, se usó el software del instrumento, el IndicoPro 6.3 (ASD inc, USA). Para el registro de los espectros de lotes de producción, se usó el software de monitorización en tiempo real del proceso Unscrambler X Process Pulse 1.1M (CAMO Analytics AS, Noruega). Además, una vez el sistema PAT estuviera en rutina, este serviría para la predicción de los CQAs y el envío de la información. Finalmente el análisis de los espectros, la calibración y la validación de los modelos se usó el Unscrambler X 10.3 (CAMO Analytics AS, Noruega).

6.3.6. Tratamiento de los datos

Para la identificación del producto se utilizó como criterio de similitud el coeficiente de correlación. El espectro de referencia se calculó promediando los espectros de 5 lotes de producción que representaban la variabilidad esperada. Los modelos cuantitativos se calcularon con el algoritmo PLS. En todos los casos se usó el promedio de los 3 espectros registrados para cada muestra (tanto de producción como de laboratorio) para poderlos relacionar con el valor de referencia. El rango espectral y pretratamiento escogidos para el cálculo se optimizaron mediante la validación cruzada y detallan en el apartado de resultados y discusión.

6.4. RESULTADOS Y DISCUSIÓN

6.4.1. Modelos predictivos

6.4.1.1. Estudio inicial

Al realizar el análisis en paralelo de un número importante de lotes de producción, por NIR y de la forma tradicional (Figura 6.4), quedó patente la desviación significativa entre los valores predichos con los modelos que estaban implementados inicialmente y los valores de referencia. En la Figura 6.4 se muestran los residuales que se obtuvieron tanto para el contenido de API como el contenido en agua. Los residuales se calcularon mediante la diferencia de los valores de referencia y los

valores predichos. Para el modelo de concentración de API, se obtuvieron residuales de hasta 5mg/g, lo que representa un 10% de error respecto al contenido nominal (50mg/g). También se observaba un sesgo negativo, esto es, los valores del NIR eran sistemáticamente superiores a los de referencia. Por otro lado, en el caso del modelo para determinar el contenido en agua, presentaba un sesgo positivo, por lo que los valores predicho con el NIR eran sistemáticamente inferiores a los de referencia.

Al realizar un test t de muestras emparejadas, se confirmó lo que se intuía visualmente. Las diferencias entre los valores predichos con los modelos implementados y los de referencia eran estadísticamente significativas, dado que los valores absolutos calculados de t (9.6 y 12.3 para el contenido de API y en agua respectivamente) fueron superiores al valor crítico de t (2.0)

La primera medida correctiva que se propuso fue la de incluir nuevas muestras de producción a los modelos, dado que suele ser necesario alimentar al sistema cuando aparecen nuevas fuentes de variabilidad. Esta estrategia fracasó, en ningún momento se consiguió validar ningún modelo con la robustez suficiente que permitiera la liberación de los mismos en rutina. Esto llevó a descartar esta posibilidad y se decidió realizar modelos nuevos, preparando nuevas muestras de laboratorio y desarrollar los modelos en las mismas instalaciones donde se iban a usar.

Antes de proceder al desarrollo de los modelos se realizó un análisis de riesgos (Tabla 6.6) con el que se identificaron todas las posibles variables que podrían ser susceptibles de afectar al espectro NIR, y en consecuencia, a la capacidad predictora de los modelos.

Tabla 6.6 Variables que pueden tener un efecto en los valores predichos con NIR y su control de riesgo.

Variable	Efecto	Control del riesgo
Condiciones ambientales	Medida errónea debido a las condiciones ambientales	Incluir en el conjunto de calibración todas las condiciones ambientales posibles
Suciedad en la interfase de muestreo	Resultados erróneos debido a que queda polvo en la ventana de zafiro	Validación del sistema de limpieza
Calidad óptica de la ventana de zafiro	Medidas erróneas	La ventana se protegerá de manera adecuada después de las medidas
Propiedades de la muestra	Resultados erróneos debido a los cambios en las propiedades físicas o químicas de las muestras	- Robustez del modelo - Evaluación del modelo y, si es necesario, modificarlo para incluir la nueva variabilidad
Variaciones locales en la concentración o la humedad	Medida no representativa de la muestra	- Proceso de mezcla validado - Se realizan tres replicados por muestra
Cambios en los consumibles del instrumento NIR	Resultados erróneos	- Robustez del modelo - Evaluación del modelo y, si es necesario, modificarlo para incluir la

Tabla 6.6 Variables que pueden tener un efecto en los valores predichos con NIR y su control de riesgo.

Variable	Efecto	Control del riesgo
		nueva variabilidad
Cambios en el proceso de producción	Resultados erróneos	Evaluación del modelo y, si es necesario, modificarlo para incluir la nueva variabilidad

Las variables del proceso o propiedades de la muestra identificadas como las que podrían afectar al espectro NIR fueron:

- El API usado en la formulación dado que hay dos fabricantes autorizados.
- La temperatura del producto analizado ya que se puede encontrar a distintas temperaturas, dependiendo del momento que se adquieran los espectros. Cuando el producto sale del equipo de granulación se encuentra a unos 70°C.
- El equipo de granulación usado debido a que hay dos procesos validados con dos equipos distintos.

Estas variables mencionadas se tuvieron presentes a la hora de escoger los lotes de producción a incluir en los conjuntos de calibración y optimización de los modelos.

6.4.1.2. Características de los espectros

Los espectros que se muestran en las figuras de esta sección están en reflectancia sobre la base de que el modelo de cuantificación de API demostró tener mejor bondad usando los espectros en estas unidades, tal como se justificará más adelante.

Los espectros adquiridos para las muestras de producción y las preparadas en el laboratorio presentaban ciertas diferencias, tal como se puede apreciar en la Figura 6.5. Las diferencias que se aprecian entre las muestras de laboratorio y las de producción se atribuyeron a la distinta

compactación que sufre la muestra. Las muestras de laboratorio tienen la presión ejercida por 100g de producto, mientras que las de producción tienen la presión de 400kg de producto. La parte del espectro con una diferencia más significativa se encuentra en el área del visible, concretamente en la banda con el máximo a 380nm. Este rango se descartó en todos los cálculos de los modelos quimiométricos, ya que es una variabilidad que no se observará en los espectros en rutina. En el resto del rango espectral fue necesario aplicar un pretratamiento que permitiera la reducción de las diferencias en las líneas bases causadas por distintos procesos de *scattering*. También se descartaron las longitudes de onda comprendidas entre 990 y 1020 nm, debido a que aparece una perturbación en la longitud de onda de 1000nm coincidiendo con la región del cambio de detector. Una vez establecidas estas restricciones, cada uno de los modelos se desarrolló combinando el rango espectral y el pretratamiento más adecuado que permitían obtener resultados exactos y robustos.

De las variables de proceso y de muestras que se identificaron durante el análisis de riesgos según lo observado en la Figura 6.6 aparentemente no tienen efecto en el espectro. Ahora bien, teniendo en cuenta que el espectro NIR puede esconder información que no se observa visualmente o a primera vista, esta variabilidad se tuvo en cuenta a la hora de escoger los lotes de producción que formarían parte del conjunto de calibración.

6.4.1.3. Modelo de identificación

La comparación del espectro puro del API y el placebo, frente al producto granulado de Nimesulide de la Figura 6.7, permite distinguir en éste último la presencia de bandas que corresponden a las más intensas del API, y que no están presentes en el placebo. Estas diferencias deberían permitir obtener un método de identificación capaz de discernir entre el producto granulado de Nimesulide y el resto.

El espectro de referencia que serviría para calcular el coeficiente de correlación fue el resultado de promediar los espectros de cinco lotes, los cuales representaban la variabilidad que se esperaba en las futuras muestras. Estos incluían productos fabricados con el API del fabricante 1 y el granulador 1, API del fabricante 2 y el granulador 1 y API del fabricante 1 y el granulador 2. La combinación API del fabricante 2 y el granulador 2 no se había producido en ningún momento, por lo que no se tuvo en cuenta en el desarrollo del modelo. El rango espectral seleccionado fue de 1020 a 2320 nm, y no se aplicó ningún pretratamiento, debido a que el software de monitorización no lo permite. Esto representó un contratiempo, ya que no era posible reducir las diferencias físicas e incrementar las diferencias químicas, por lo que cobraba vital importancia establecer un nivel de aceptación óptimo. Este *threshold* tenía que permitir identificar correctamente muestras de producción con una diversidad elevada de variabilidad y, a la vez, permitir diferenciar productos espectralmente muy similares, como es el placebo. Se estableció un nivel de aceptación de $R \geq 0.998$. La validación externa demostró la especificidad o selectividad del método y su robustez. En el gráfico de la Figura 6.9 se representan los valores de coeficiente de correlación obtenido con el producto granulado de Nimesulide, el placebo, la sacarosa y con el otro producto granulado que se manipula en el mismo bin.

En el gráfico de la izquierda de la Figura 6.9 se representan los valores con lotes del producto granulado de Nimesulide, fabricados con los dos API disponibles. De los 390 espectros usados en la validación externa, hay 8 que han resultado mal identificados (coeficiente de correlación menor al umbral establecido). Para comprender el motivo del fallo, se compararon sus espectros con el de referencia (Figura 6.8 izquierda). Los lotes correspondientes a los puntos 248, 249, 250 y 276 pertenecían a lotes con un contenido en agua por encima o muy cerca del límite de especificación interna. Esto se corrobora al comparar los espectros de los mismos con el de referencia, ya que en los primeros se observa una banda más intensa sobre 1940nm. Este pico corresponde a la banda de combinación del agua⁵. Para el resto de muestras (78, 192, 193 y 335) visualmente no se observaron diferencias significativas en los espectros. Se calculó un modelo PCA con los espectros usados para establecer el espectro de referencia e incluyendo los primeros 2 PC, los cuales explicaban el 99% de la varianza de los datos. Al proyectar las muestras del producto granulado de Nimesulide correspondientes a los lotes de validación externa, los mismos espectros que habían dado un coeficiente de correlación por debajo del nivel de aceptación establecido, también aparecieron fuera de los límites de Hotelling T^2 y/o residuales Q. Se concluyó que dichos espectros

presentaban una variabilidad distinta a la considerada en el cálculo de la referencia y se asumió este porcentaje de falsos negativos.

El gráfico de la derecha de la Figura 6.9 demuestra la selectividad del método. Éste es capaz de descartar los productos parecidos (sacarosa como excipiente mayoritario y placebo granulado) y el producto granulado que se procesa en la misma instalación.

6.4.1.4. Modelos PLS

Para la cuantificación de la concentración de API y el contenido en agua en el producto granulado de Nimesulide se calcularon dos modelos PLS que permitieron relacionar los espectros NIR con los valores de referencia. Se ensayaron distintas combinaciones de pretratamientos (derivada Savitky-Golay, SNV, MSC) y rangos espectrales. Se escogieron los que presentaron mejores resultados haciendo uso de la validación cruzada. En la Tabla 6.7 se presentan las figuras de mérito de ambos modelos.

Tabla 6.7 Figuras de mérito de los modelos de calibración PLS.

Características	PLS cuantificación API	PLS contenido en agua
Número de muestras (Laboratorio + Producción)	56 (45+11)	43 (35+8)
Modo del espectro	Reflectancia	Absorbancia
Pretratamiento espectral	1ª Derivada SG (polinomio orden 2, ventana móvil 15 puntos)	1ª Derivada SG (polinomio orden 2, ventana móvil 17 puntos)
Rango espectral (nm)	1014-1276; 1458-2306	729-1968
Rango de calibración	34.1-67.3 mg/g	0.12-1.67 %
Número de factores PLS	5	4
Varianza explicada (Y) (%)	98.9	95.4
Regresión Y_{ref} vs. Y_{NIR}		
Pendiente	0.989	0.954
Ordenada en el origen	0.55	0.02
RMSEC/RMSECV	0.88/1.15 (mg/g)	0.08/0.10%

Cuantificación de API

Para desarrollar el modelo PLS capaz de relacionar los espectros NIR con la concentración de API en el producto granulado de Nimesulide se incluyeron 57 muestras en el conjunto de calibración. De ellas, 47 se habían preparado en el laboratorio para ampliar el rango de concentración y las otras 10 pertenecían a muestras de producción. La elección de estas últimas se realizó teniendo en cuenta: (1) el conocimiento previo de la variabilidad que contenían las muestras debido al proceso y (2) la elección de las que presentaban mayor variabilidad espectral. Para este último supuesto, se escogieron usando el gráfico de *scores* una vez calculado el PCA, de manera que las muestras de

calibración representaran la variabilidad total. Los espectros de todas las muestras de calibración están representados en la Figura 6.10. El rango espectral seleccionado para el cálculo fue 1014-1276 nm y 1458-2306 nm después de aplicar la primera derivada Savitky-Golay de orden 2 y ventana móvil de 15 puntos. El rango de 1277-1457 nm se descartó debido a que la variabilidad que presenta esta banda no estaba relacionada con la concentración de API y los modelos obtenidos resultaron tener un RMSECV elevado.

El gráfico de *scores* del modelo PLS calculado (Figura 6.11) muestra como el primer factor no explica exclusivamente la diferencia entre muestras por su concentración (solo el 27%), sino que también se ve influenciado por el tipo de muestra analizada (de producción y de laboratorio). No se encontró una combinación de pretratamiento y rango espectral capaz de eliminar por completo la diferencia en el efecto *scattering* producido por la disparidad de presiones (o compresiones) de la muestra, si bien el modelo ha demostrado tener un poder de predicción suficiente como demuestra la validación externa del mismo.

Un hecho sorprendente en la calibración de este modelo de cuantificación de API fue que los modelos calculados con el espectro en modo absorbancia daban unos residuales 2 veces superiores a los obtenidos si se calculaba con los espectros en reflectancia. En general, cuando se usa el espectro NIR de reflectancia difusa para la cuantificación de un analito suele cumplirse la Ley de Beer, la cual relaciona linealmente la absorbancia con la concentración. Aunque se dé alguna pequeña desviación de la linealidad el PLS es capaz de resolverlo⁶. Ahora bien, en el rango de concentraciones de API en el que se trabaja, la desviación de la linealidad es tan elevada, que no es posible corregirlo con el PLS. En estas situaciones la bibliografía consultada propone el uso de la ecuación de Kubelka-Munk^{7,8}. Pero en este caso concreto se ha demostrado que la reflectancia presenta una relación lineal con la concentración de API en el rango de estudio, por lo que se ha escogido para el cálculo del modelo PLS.

Las figuras de mérito de la validación externa del modelo de cuantificación (Tabla 6.8) demostraron la habilidad del mismo para predecir la cantidad de API presente en nuevas muestras del producto granulado de Nimesulide.

Cuantificación contenido en agua

La determinación del contenido en agua en un producto mediante NIR es una de las aplicaciones más extendidas debido a que el coeficiente de extinción del agua es muy elevado⁹. Para la calibración del modelo PLS que relacionara los espectros NIR con el contenido se emplearon un total de 43 muestras, de las cuales 35 provenían de muestras preparadas en el laboratorio y 8 de producción. Estas últimas se eligieron siguiendo los mismos criterios que los comentados anteriormente, las que contenían variabilidad del proceso y variabilidad espectral. Inicialmente se disponía de 37 muestras de laboratorio, pero al calcular el modelo se detectó que las muestras que contenían más de un 1.80% de agua aparecían como *outliers*, distorsionando el modelo de tal manera que se perdía la linealidad del mismo. Se decidió descartar las dos muestras que se prepararon con un contenido alrededor del 2.00%, considerándose que era suficiente ampliar el rango hasta 1.67% dado que las especificaciones internas permiten un máximo del 0.50%.

El rango espectral incluido en el cálculo, después de aplicar una primera derivada al espectro en absorbancia fue 729-1968 nm (Figura 6.13). En esta región aparecen los cinco máximos de absorción del agua, 760, 970, 1190, 1450 y 1940 nm¹⁰.

En el gráfico de *loadings* de los cuatro factores incluidos en el modelo, en la Figura 6.12, se observan picos correspondientes a las bandas del agua que se han mencionado, por lo que se puede deducir que todos ellos explican una parte de la variabilidad de los datos debido a las diferencias de contenidos de agua.

Las figuras de mérito de la validación externa del modelo de cuantificación (Tabla 6.8) demostraron la habilidad del mismo para predecir el contenido en agua presente en nuevas muestras de producción.

Implementación de un sistema PAT en la industria farmacéutica

Tabla 6.8 Figuras de mérito obtenidos con la validación externa de los modelos PLS de cuantificación de API y del contenido en agua.

Parámetros de validación		PLS cuantificación API	PLS contenido en agua
Especificidad	Coefficiente de correlación método identificación	≥ 0.998	≥ 0.998
	Número de muestras	15	12
Linealidad y rango	Rango	35.1-65.9 mg/g	0.21-1.23 %
	Varianza explicada Y	99.3%	96.2%
	Pendiente \pm IC ($\alpha=0.05$) incluye 1	0.96 ± 0.04	1.07 ± 0.10
	Ordenada en el origen \pm IC ($\alpha=0.05$) incluye 0	2.20 ± 2.23	-0.04 ± 0.09
	$R^2 > 0.95$	0.99	0.99
	Aleatoriedad de los residuales	Cumple (Figura 6.14)	Cumple (Figura 6.14)
	Número de muestras	15	12
Exactitud	RMSEP	0.70mg/g	0.04%
	Test t de muestras emparejadas $t_{exp} < t_{calc}$	$t_{exp} = -0.12$ $t_{calc} = 2.14$	$t_{exp} = 0.74$ $t_{calc} = 2.20$
	RSD de seis medidas de la misma muestra $\leq 2\%$	0.50	0.60
Precisión intermedia	ANOVA 2 factores (3 días y 2 analistas) no tienen efecto significativo	Analista: $\rho = 0.586$ Día: $\rho = 0.075$	Analista: $\rho = 0.113$ Día: $\rho = 0.364$
Robustez	Número de muestras	117	125
	Test t de muestras emparejadas $t_{exp} < t_{calc}$	$t_{exp} = -1.19$ $t_{calc} = 1.98$	$t_{exp} = -1.28$ $t_{calc} = 1.98$

6.4.1.5. Monitorización y mantenimiento de los modelos

Con el paso del tiempo los modelos NIR pueden verse afectados por cambios en los materiales de partida, cambios de parámetros de proceso, cambios en el instrumento o por factores que no se

tuvieron en cuenta durante el desarrollo del modelo. Es imprescindible establecer una estrategia que permita monitorizar el comportamiento de los modelos y mantenerlos. Hay diversos métodos reportados en la literatura para monitorizar los modelos en rutina y detectar cuándo empiezan a fallar¹¹⁻¹³. En este trabajo se han establecido varias medidas para poder mantener los modelos que se han desarrollado en las secciones anteriores.

Como método de monitorización principal se propusieron los gráficos de control aplicados a los residuales que se obtienen entre los valores predichos y los valores de referencia. Esta herramienta se caracteriza por ser de sencilla implementación y manejo, por lo que lo puede usar personal no experto en quimiometría. A partir de los residuales que se obtuvieron con las predicciones de las muestras empleadas para evaluar la robustez de los modelos, se definieron la línea central, los límites de alerta y los de acción (\bar{X} , $\pm 2\sigma$ y $\pm 3\sigma$ respectivamente) de los gráficos de control que se usarían para la monitorización del funcionamiento de los modelos NIR.

Tabla 6.9 Parámetros de los gráficos de control calculados a partir de las muestras de validación externa.

	Valores para el PLS de cuantificación de API	Valores para el PLS de la cuantificación de contenido en agua
Límite superior de acción	2.50	0,17
Límite superior de alerta	1.70	0,12
Línea central	0.09	0,01
Límite inferior de alerta	-1.52	-0,10
Límite inferior de acción	-2.33	-0,16

De manera rutinaria, cada diez lotes de producción, uno se analizará por ambos métodos, NIR y referencia. Los residuales obtenidos se representarán en los gráficos de control con los parámetros que se han definido en la Tabla 6.9. Se aplicarán las reglas de Nelson para detectar puntos fuera de control, tendencia o desplazamientos. Si en algún momento se detecta un patrón el modelo se evaluará de manera minuciosa.

En el caso de detectar patrones se realizará la segunda etapa de la monitorización, el uso de diagnósticos internos del modelo. Estos son los valores de influencia de Hotelling T^2 y los residuales Q que se establecieron. Con el conjunto de calibración empleado para el cálculo de cada modelo PLS se establecieron los límites con un nivel de confianza del 95% para cada uno de los parámetros (modelo API: T^2 de 13.15 y Q de $1.2 \cdot 10^{-6}$; modelo agua: T^2 de 11.41 y Q de $2.6 \cdot 10^{-8}$). Si al proyectar los espectros de las muestras que presenten un patrón se obtienen unos valores de Hotelling T^2 inusualmente elevados, significará que se encuentra a una distancia elevada del centro del modelo, pero en el mismo plano¹². En este caso, la predicción de la muestra se hace por extrapolación, por lo que estas pueden llevar a una disminución moderada de la exactitud de los valores predichos¹¹. Por otro lado, si las muestras presentan un residual Q elevado, significará que

los espectros presentan una variación ortogonal al plano del modelo, la cual no queda explicada por el modelo¹². En este caso, los problemas de inexactitud en los valores predichos son mayores¹¹. En ambos casos, si con este diagnóstico se confirma que la causa raíz del problema es que los espectros presentan una variabilidad nueva y aceptada respecto a las muestras usadas durante la calibración, el modelo se deberá actualizar. Para ello se incluirá esta nueva variabilidad en el conjunto de calibración y una vez calculado el modelo se revalidará por completo. Por lo contrario, si la causa raíz es un problema inaceptable (por ejemplo mal funcionamiento del instrumento, deriva del proceso de producción...) el modelo no se actualizará sino que se afrontará el problema que causa dicha desviación.

Otra medida considerada de monitorización del modelo es que siempre que se realicen cambios especificados en el análisis de riesgos de la Tabla 6.6 los cuales pueden afectar al espectro y, en consecuencia, a la capacidad de predicción de los modelos, se evaluará el impacto del cambio. Para ello se analizarán un seguido de lotes por ambos métodos, NIR y referencia, y se determinará la necesidad de actualizar el modelo.

Por último, siempre que haya un valor fuera de especificación (OOS) el lote implicado se analizará por el método de referencia para confirmarlo. Si no es confirmado, se estudiarán las causas raíces y, si fuera necesario, se recalculará el modelo.

6.4.2. Implementación del sistema PAT

La implementación industrial del sistema PAT ha consistido en un conjunto de tareas que se hacen imprescindibles para que se pueda usar de manera rutinaria. El esquema general de todos los elementos implicados en el análisis de un lote por NIR, así como las comunicaciones, se presenta en la Figura 6.15.

El proceso que se sigue cuando se realiza una nueva orden de trabajo (OT) para la granulación y la mezcla del producto, empieza con la generación de un lote asociado a la OT en el sistema de planificación de recursos empresariales (ERP por sus siglas en inglés *Enterprise Resource Planning*). Además, de manera automática, se genera una petición de análisis en el sistema de gestión de información de laboratorio (LIMS) asociada a la OT. Hasta aquí no hay ninguna diferencia con el modo tradicional de proceder. La parte nueva es que es necesario crear una ficha relacionada con el lote y OT del producto a analizar en el software de control en proceso (IPC por sus siglas en inglés *In Process Control*). Cuando el bin del lote en cuestión se mezcla, en el panel mezclador se introducen los datos de la ficha IPC. Automáticamente se envían las tramas del protocolo TCP/IP (protocolo que permite la comunicación entre máquinas) entre el ordenador (PC del inglés *Personal Computer*), el controlador lógico programable del aplicativo IPC (PLC del inglés *Programmable Logic Controller*) y el panel del equipo del mezclador de depósito bin. Esto permite que desde el panel del mezclador se puedan dar órdenes de adquisición de espectros – tanto de la referencia como de la muestra- y estas se envíen, a través del PLC IPC, al software para la monitorización en tiempo real. Éste último, cuando recibe la orden, es el que controla el NIR. Cuando el instrumento NIR ha adquirido los espectros, el software de monitorización predice los CQAs con los modelos quimiométricos integrados. En este momento, se genera una nueva trama del protocolo TCP/IP y los valores se envían al PLC IPC. Cuando éste último ha recibido tres valores para cada CQA (recordemos que un lote se analiza por triplicado) se terminan las tramas. Desde el software IPC se recuperan los datos generados y se introducen en el LIMS. El departamento de control de calidad revisa los resultados y dictamina el lote. Si es conforme, el producto se libera a través del ERP para que pueda pasar a la siguiente operación unitaria, la dosificación. El último elemento clave en este proceso es el sistema de monitorización y distribución de la información (SMDI), el cual permite el registro de todas las operaciones realizadas durante el proceso de mezcla y el análisis por NIR. Éste tiene funciones de *Audit Trail*.

Para poder implementar todo este complejo sistema, se tuvieron que realizar una serie de cualificaciones y comprobaciones para corroborar y demostrar documentalmente el buen funcionamiento de todas las partes implicadas en el sistema PAT según las normas GMP y el cumplimiento de la integridad de datos según la 21CFR parte 11 (Tabla 6.10). Además se desarrollaron las instrucciones de uso pertinentes. El resto de cualificaciones no especificadas, ya se habían realizado en la compañía previamente.

Tabla 6.10 Cualificaciones realizadas al instrumento NIR, al software de monitorización en tiempo real y al sistema PAT en general.

Elemento	Acciones realizadas
Instrumento NIR y Software para la	<u>IQ:</u> - Caracterizar software y hardware

Implementación de un sistema PAT en la industria farmacéutica

Tabla 6.10 Cualificaciones realizadas al instrumento NIR, al software de monitorización en tiempo real y al sistema PAT en general.

Elemento	Acciones realizadas
monitorización en tiempo real	<ul style="list-style-type: none"> - Caracterizar la configuración - Corroborar que los componentes están instalados según el suministrados - Corroborar la disponibilidad de documentación descriptiva (manuales, características...) - Comprobar la existencia de medios para recuperar el sistema en caso de fallo (copia imagen, backup...) - Comprobar la existencia de PNT de gobernanza (backup, controles de cambio, incidencias...) <p><u>OQ:</u></p> <ul style="list-style-type: none"> - Comprobar las funciones básicas tanto del instrumento como del software (Audit trail, accesos, informes, backup, interfase con otros sistemas...) - Mirar las funcionalidades contra los requerimientos <p><u>PQ:</u></p> <ul style="list-style-type: none"> - Comprobar el correcto funcionamiento en condiciones reales - Comprobar que hay los PNT de uso - Comprobar que se ha realizado formación
Sistema PAT	<p><u>PQ:</u></p> <ul style="list-style-type: none"> - Comprobar el correcto funcionamiento en condiciones reales - Comprobar que hay los PNT de uso - Comprobar que se ha realizado formación

Finalmente, se analizaron 40 lotes en paralelo –determinación por el método NIR y por el de referencia- para poner a prueba el sistema PAT antes de liberarlo en rutina. Este momento se aprovechó para formar a personal involucrado en las medidas.

Los gráficos de control obtenidos con las cuarenta muestras analizadas en paralelo se presentan en la Figura 6.16 y la Figura 6.17. Antes de pasar a comentar los gráficos, cabe mencionar que todos los valores predichos (individuales y promedio de cada lote) a partir del espectro NIR fueron conformes, o lo que es lo mismo, estuvieron dentro de especificaciones.

En el gráfico de control para la concentración de API solo se observaron dos patrones de Nelson. El primero corresponde a la muestra señalada con un punto rojo, el valor del cual está por fuera de -3σ . El otro patrón de Nelson es el que corresponde a las muestras 25 y 26, dos puntos consecutivos están por debajo de -2σ . Las tres muestras correspondían al producto fabricado con el API homologado 2. Haciendo una evaluación de los valores obtenidos en esta fase comparativa y durante el análisis de la robustez del modelo se observó cómo, de manera sistemática, cuando el producto analizado se había formulado con el API del fabricante 2 los valores predichos eran menores que los de referencia. La conclusión a la que se llegó es que un único modelo no podía modelar de manera satisfactoria la variabilidad que presentaban ambos productos por lo que se realizarán dos modelos PLS separados para cada uno de los API.

Por otro lado, el gráfico de control obtenido para el modelo PLS capaz de predecir el contenido de agua, mostró un ligero sesgo negativo. Esto es, los valores de referencia eran de manera mayoritaria superiores a los predichos con el NIR. Este comportamiento ya se intuyó evaluando la linealidad del modelo durante la validación externa, al comprobar la aleatoriedad de los residuales (Figura 6.14). Ahora bien, una mejora en este sentido sería muy complicada de obtener. La causa raíz se encuentra en la manera de obtener los valores por ambos métodos. En el caso del análisis por NIR, se realiza a través de la ventana de zafiro sin que, en ningún momento, el producto esté en contacto con el ambiente. En el caso de la pérdida por desecación, el método de referencia, el producto está en contacto con el ambiente antes de realizar el análisis (durante el muestreo, durante la pesada...), por lo que siempre puede absorber humedad. Además, este fenómeno de absorción puede variar cíclicamente siendo dependiente de la estacionalidad. Como el modelo ha demostrado ser robusto con un gran número de lotes que cubren todas las estacionalidades y ser capaz de detectar fuera de especificaciones internas o lotes que están cerca de ellas, se considera que el modelo es funcional para el análisis.

6.5. CONCLUSIONES

En este trabajo se han desarrollado y validado de manera satisfactoria tres nuevos métodos NIR que permiten identificar y predecir la riqueza de API y el contenido de agua del producto granulado de Nimesulide. El método de identificación empleado ha sido el coeficiente de correlación. En el caso de los modelos de cuantificación se ha usado como algoritmo el PLS. Para el cálculo de la riqueza de API se ha observado que, en el rango de estudio, la reflectancia muestra un comportamiento lineal con la concentración, no siendo así con la absorbancia tal como marca la ley de Beer. También se ha establecido el mecanismo de monitorización y mantenimiento de los modelos, mediante gráficos de control en primera instancia, y diagnósticos de los modelos en segunda.

Por otro lado, también ha sido posible concluir la implementación de un sistema PAT de forma industrial. Para ello ha sido necesario participar en todas las cualificaciones pertinentes que permitieran el uso del instrumento, software y sistema de monitorización en rutina. La implementación de este sistema PAT ha de permitir disminuir el tiempo de stock de granel del medicamento, dado que el cuello de botella es la espera de las horas o los días necesarios para obtener los resultados desde el laboratorio de control de calidad, provocando que se retrase el inicio de la siguiente operación unitaria, la dosificación. Esta mejora productiva se prevé que comporte una disminución relevante del coste de fabricación de la familia de medicamentos.

6.6. BIBLIOGRAFÍA

1. Bakeev, K. A. *Process Analytical Technology: Spectroscopic Tools and Implementation Strategies for the Chemical and Pharmaceutical Industries*. (Wiley, 2010).
2. Rosas, J. G., Blanco, M., González, J. M. & Alcalà, M. Real-time determination of critical quality attributes using near-infrared spectroscopy: A contribution for Process Analytical Technology (PAT). *Talanta* **97**, 163–170 (2012).
3. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Anexo 15 Cualificación y validación. (2015).
4. European Pharmacopoeia. Ph. Eur. 9.0 Chapter 2.2.40. Near-infrared spectroscopy. (2017).
5. Roggo, Y. *et al.* A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J. Pharm. Biomed. Anal.* **44**, 683–700 (2007).
6. Griffiths, P. R. Letter: Practical consequences of math pre-treatment of near infrared reflectance data: $\log(1/R)$ vs $F(R)$. *NIR Publ.* **62**, 60–62 (1995).
7. Elliott, S. & Evans, J. A 3-year review of new psychoactive substances in casework. *Forensic Sci. Int.* **243C**, 55–60 (2014).
8. Burns, D. A. & Ciurczak, E. W. *Handbook of Near Infrared Analysis*. (CRC Press Taylor and Francis, 2008).
9. Ciurczak, E. W. & Igne, B. *Pharmaceutical and Medical Applications of Near-Infrared Spectroscopy*. (CRC Press, 2002).
10. Reich, G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications.

- Adv. Drug Deliv. Rev.* **57**, 1109–1143 (2005).
11. Wise, B. M. & Roginski, R. T. A calibration model maintenance roadmap. *IFAC-PapersOnLine* **48**, 260–265 (2015).
 12. Cogdill, R. P., Anderson, C. A. & Drennen, J. K. Process Analytical Technology Case Study , Part III : Calibration Monitoring and Transfer. *AAPS PharmSciTech* **6**, 284–297 (2005).
 13. Nikzad-langerodi, R. *et al.* Calibration model maintenance in melamine resin production : Integrating drift detection , smart sample selection and model adaptation. *Anal. Chim. Acta* **1013**, 1–12 (2018).

7. ESTUDIO DE VIABILIDAD DE UN MÉTODO NIR PARA IMPLEMENTAR EN LA ETAPA DE DOSIFICACIÓN

7.1. INTRODUCCIÓN

El contenido de API y su coeficiente de variación de los productos farmacéuticos son dos parámetros críticos de calidad que deben ser determinados. La implementación de un sistema PAT que permitiera determinar dichos parámetros durante la dosificación de productos granulados, polvos o comprimidos contribuiría a mejorar el conocimiento del proceso y, además, los lotes se podrían liberar en tiempo real. En el caso de los comprimidos, también se podría demostrar la uniformidad de contenido a partir de un número elevado de muestras, de 100 a 1000, siguiendo las directrices de la monografía 2.9.47 de la Farmacopea Europea¹, con unos valores de aceptación adaptados a esta nueva realidad. Estos criterios de aceptación tienen un margen más amplio que los aplicados en el análisis de uniformidad tradicional, en el que se analizan 10 o 30 comprimidos².

La elección del instrumento NIR que pueda instalarse en el proceso de producción y permita el análisis de las muestras en movimiento, es uno de los primeros pasos que se debe realizar para poder llevar a cabo la implementación de un sistema PAT. En la actualidad existen varias tecnologías que lo permiten. La gran mayoría consisten en instrumentos con los que se obtiene un espectro, ya sea el que proviene de un detector o del promedio de varios detectores, con cada medida realizada (*single-point*)³. Pero también existen equipos NIR con sonda multipunto, como la espectroscopia espacialmente resuelta (SRS por sus siglas en inglés *Spatially Resolved Spectroscopy*). Con ellos se obtienen un número determinado de espectros independientes que dan información espacial de la muestra^{3,4}. Están formadas por una fuente de iluminación y varios canales (detectores) situados a distintas posiciones respecto a la fuente. Dependiendo de la distancia que los separa de la fuente la información que se obtiene varía. Los espectros adquiridos con los que están más cerca están caracterizados por el fenómeno de *scattering*, mientras que a medida que la distancia entre el detector y la fuente aumenta, la información espectral corresponde a capas más profundas de la muestra³. Estos equipos se encuentran a medio camino entre los NIR clásicos y las cámaras hiperespectrales, por lo que permiten desarrollar métodos para: (1) la determinación de parámetros de calidad de la muestra analizada, empleando el espectro promedio de la muestra, y (2) obtener información sobre la homogeneidad de la misma si los espectros se usan de forma individual³.

El objetivo de este trabajo consistió en poner a prueba las capacidades de un instrumento NIR, que registra de manera simultánea diferentes espectros de una muestra, para dar respuesta relacionada con la concentración de API de un producto farmacéutico que está en movimiento. Esta prueba de

viabilidad forma parte de la etapa inicial de un proyecto PAT, con el cual se pretende instalar un equipo NIR e implementar un método en las etapas de dosificación que permita liberar en tiempo real el producto farmacéutico que se está analizando.

7.2. MÉTODO EXPERIMENTAL

7.2.1. Muestras

Los productos elegidos para este estudio fueron los siguientes: un producto en polvo, de mezcla directa con tendencia a la segregación, el cual tiene un 5.7% de API; dos productos granulados que contienen un 5.0 y un 1.0 % de API; y unos comprimidos recubiertos con un 10.0% de API.

La estrategia seguida para preparar las muestras a distintos niveles de concentración (del 70 a 130 % del contenido declarado) fue la de sub-dosificar y sobre-dosificar muestras de producción añadiendo placebo o API respectivamente. Normalmente, esta estrategia requiere preparar distintos placebos de acuerdo a un DoE para variar la cantidad de los distintos excipientes un $\pm 5\%$ del contenido declarado, y evitar la correlación entre ellos y el API. En este caso, se decidió preparar un solo placebo con los valores nominales de excipiente de la formulación, evitando introducir variabilidad extra dado que el objetivo del estudio era probar la habilidad del instrumento a detectar cambios y no desarrollar un modelo robusto que se pudiera usar en rutina.

A partir de 3 lotes distintos de producción para cada producto se prepararon las muestras con los correspondientes niveles de concentración de API cada uno. Así, para cada producto se obtuvieron 15 muestras, 3 procedentes de las muestras de producción y 12 preparadas en el laboratorio. Para la homogeneización de las muestras de laboratorio se empleó un mezclador en forma de V de la planta piloto. Los comprimidos se produjeron en la planta piloto mediante el uso de una comprimidora y con los mismos punzones de fábrica, permitiendo obtener comprimidos con la misma forma y características físicas (dureza, peso y grosor) que los de producción.

Las concentraciones gravimétricas de las muestras se utilizaron como valor de referencia para desarrollar los modelos cuantitativos.

7.2.2. Adquisición de espectros

El estudio se realizó con el instrumento Hy-Trinity NIR (Indatech, Francia), el cual cubre el rango espectral de 900 a 1700 nm. El instrumento estaba equipado con una sonda de no contacto con varias fibras ópticas. La configuración de la misma se presenta en el esquema de la Figura 7.1. Las fibras ópticas activadas durante la adquisición de los espectros fueron las que aparecen en azul en el esquema. Un *scan* contenía 8 espectros. Cabe mencionar que la distancia desde el punto de iluminación a las fibras 42, 43 y 44 es la misma que desde el mismo punto de iluminación a las

fibras 42b, 43b y 44b. Esta sonda estaba colocada en una cinta que permitía simular el proceso real de análisis, donde el producto pasaba por debajo de ella con una velocidad fijada a 200mm/s en todos los experimentos. La distancia de la sonda a las muestras se optimizó manualmente de manera que la presencia de ruido en los espectros quedara minimizada.

El registro de las muestras en polvo o granuladas se realizó colocando 10g en el soporte y procurando que la superficie quedara plana. Esto se hizo para evitar la variabilidad espectral debida a la diferencia de distancias entre la sonda y la muestra. El soporte se colocaba en la cinta en movimiento, de manera que la muestra pasaba por debajo de la sonda NIR. Los espectros se registraron con un tiempo de integración de 2ms. En el caso de los comprimidos, se registraron espectros de dos maneras. En la primera se usaron las mismas condiciones que las mencionadas anteriormente, pero los espectros obtenidos presentaban ruido. Para mejorar la relación S/N se disminuyó la distancia entre la sonda y el soporte. Para este segundo registro se definió un tiempo de integración de 4ms. Este aumento responde a que la saturación disminuye al disminuir la distancia entre la sonda y la muestra. Para mantener un nivel de saturación constante (80%) fue necesario aumentar el tiempo de integración. Para el análisis de cada una de las muestras se colocaron siete comprimidos en el soporte, dejando una distancia de medio centímetro entre ellos. A fin de evaluar la repetitividad de la medida se analizaron 6 veces 7 comprimidos de los tres lotes de producción.

Al inicio y final del proceso experimental se registró el espectro de referencia con un *spectralon* para verificar que no había habido cambios durante el tiempo que duró el análisis de todas las muestras. Estos mismos espectros de referencia sirvieron, más adelante, para transformar los espectros originales, los cuales se obtenían en unidades de cuenta, a unidades de absorbancia.

Para evitar tener información innecesaria, todos los espectros se adquirieron mediante lo que se conoce como *Trigger mode*. Este permite que solo se almacene información cuando el software

detecta que el producto está pasando por debajo de la sonda. El criterio por el cual se activaba el almacenaje se definió antes de realizar el análisis de todas las muestras. Para ello, se analizó de manera continua una muestra de producción de cada uno de los productos objetos de estudio, obteniendo los espectros de cuando la muestra pasaba por el NIR y de cuando no pasaba. De los datos obtenidos se definió una región óptima de longitudes de onda e intensidades donde se confirmaba que aparecía señal solo cuando la muestra pasaba por la sonda. Durante el análisis en rutina, cuando la fibra 22 de la Figura 7.1 detectaba señal en la ventana definida, daba orden al software de que almacenara los espectros generados. De este modo, la señal de la fibra 22 no se tendría en cuenta en el desarrollo de los modelos, ya que su único propósito era el que se ha definido.

7.2.3. Tratamiento de los datos

Se emplearon dos programas para el desarrollo de los resultados, el MATLAB v7.11.0 (The MathWorks, USA) y el Solo 8.1 (Eigenvector, USA). El primero sirvió para transformar los espectros en crudo a una matriz manejable para analizar. El segundo se utilizó para aplicar pretratamiento a los espectros, realizar exploración de datos y desarrollar los modelos cuantitativos con el algoritmo PLS.

El equipo proporcionaba una matriz de datos compleja. Para una muestra, se disponía de n medidas (del orden de las decenas en el caso de cada comprimido y del orden de los centenares en el caso de cada muestra de granulado o polvo). Cada medida contenía la concatenación de 8 espectros, esto es, un espectro por cada fibra óptica activada (matriz $n \times 8$). Además, en el caso de los comprimidos recubiertos, cada muestra consistía en siete comprimidos (matriz de $7 \times n \times 8$).

La preparación de las matrices para realizar los análisis pertinentes se llevó a cabo tal como se detalla en la Figura 7.2. El primer paso correspondió en eliminar aquellos espectros que contenían información distorsionada de cada muestra. Se prepararon matrices separadas seleccionando fibras individualmente, o conjuntos de ellas, con el fin de obtener información de utilidad. Una vez elegidas, era necesario transformar las intensidades que se encontraban en unidades de cuenta a unidades de absorbancia. Con este fin se promediaron los dos espectros de referencia registrados y se aplicó la ecuación $-\log(\text{espectro muestra}/\text{espectro promedio de referencia})$. Para desarrollar el modelo de cuantificación del API se utilizó el espectro promedio de las fibras 42, 43, 44, 42b, 43b y 44b.

7.3. RESULTADOS Y DISCUSIÓN

7.3.1. Justificación de la elección de los productos analizados

Estudio de viabilidad de un método NIR para implementar en la etapa de dosificación

Como se ha mencionado en la introducción, el objetivo del proyecto es implementar un método NIR que permita determinar de manera *in-line* la concentración de API de distintos productos farmacéuticos durante la etapa de dosificación de los mismos, lo cual permitiría realizar una liberación del producto en tiempo real. Para ello, se realizó un estudio de mercado con el que se identificaron distintas tecnologías disponibles que permitirían alcanzar dicho objetivo. Una de estas

tecnologías fue la que se presenta en este trabajo, un equipo NIR multipunto con tecnología SRS, que permitiría obtener información espectral y espacial de la muestra analizada⁴.

El estudio de viabilidad se diseñó de manera que se obtuvieran conclusiones sobre la adecuación de esta tecnología en la realidad de la planta de producción. Sobre la base de este objetivo, se escogieron los cuatro productos farmacéuticos empleados en este trabajo, la justificación de la cual se detalla en la Tabla 7.1.

Tabla 7.1 Justificación de la elección de los productos farmacéuticos para realizar el estudio de viabilidad.

Producto farmacéutico	Razonamiento de la elección
Granulado con un 5.0% de API	Producto granulado de mayor volumen de producción en la planta.
Granulado con un 1.0% de API	Producto granulado de volumen importante de producción con una concentración de API que se encuentra cerca del límite de detección de la técnica NIR.
Polvo con un 5.7% de API	Producto granulado con un volumen menor de producción, pero que presenta tendencia a la segregación.
Comprimido recubierto con un 10.0% de API	Producto con mayor volumen de producción de la planta.

Una de las limitaciones, a priori, más remarcables de este equipo se encontró en el rango espectral que abarca -de 900 a 1700 nm- dado que las bandas más intensas del NIR corresponden a las bandas de combinación y al primer sobretono, las cuales suelen aparecer por encima de 1400nm^{5,6}. Se realizó una comprobación previa de los espectros de los cuatro productos en granel y sus correspondientes API y placebos (Figura 7.3), los cuales se adquirieron con un equipo FT-NIR estándar de sobremesa. La comparación de los espectros obtenidos sirvió para confirmar que en todos los casos había presencia de bandas correspondientes al API en entre 900 y 1700 nm y, además, estas no coincidían con las del placebo. El rango de trabajo podría ser suficiente para observar cambios espectrales debidos a la distinta concentración.

7.3.2. Exploración de los datos

Los espectros concatenados de la Figura 7.4 corresponden al producto granulado con un 5.0% de API, y ponen de manifiesto que las fibras de la derecha (42b, 43b y 44b) presentaban una variabilidad mayor que las de la izquierda. Los espectros con menor intensidad correspondían a las primeras lecturas de la muestra. Este fenómeno se repetía en todos los análisis realizados a los productos en polvo y granulados y está relacionado con la mayor distancia que había entre el producto y la sonda, dado que en la parte del principio del soporte no estaba completamente lleno

Estudio de viabilidad de un método NIR para implementar en la etapa de dosificación

de muestra. Para no introducir esta variabilidad, la cual no estaba relacionada con la muestra, en el momento de realizar los modelos se descartaron los primeros 15 espectros.

A diferencia de los productos granulados o en polvo, en los comprimidos recubiertos no se daba este fenómeno. Pero también fue necesario descartar algunos espectros del inicio y del final de cada comprimido por dos razones. La primera era igualar el número de espectros de cada uno de los comprimidos que formaban parte de una misma muestra. Cuando se analizaban los 7 comprimidos que conformaban una muestra, el número total de espectros adquiridos en cada uno de ellos podía variar ligeramente. El análisis de los comprimidos lacados de una muestra de producción se muestra en la Figura 7.5. En él se puede comprobar que el número de espectros varía entre 13 y 14.

Esto depende de la posición relativa a la sonda de cada uno de los comprimidos en el momento en que se empiezan a adquirir los espectros, ya que el tiempo de integración es de 4ms. La segunda razón era eliminar espectros que contenían información especular. Los primeros espectros de un comprimido correspondían al borde del mismo. Dado que el borde no es plano, sino que tiene cierta curvatura, la posición del detector no era la óptima y el espectro obtenido era prácticamente plano, ya que mayoritariamente se detecta la radiación correspondiente a la reflectancia especular. En la

Figura 7.6 se representan las 13 medidas que conforman un comprimido, y se demuestra que el primer espectro es prácticamente plano, y el segundo, contiene cierta información especular. Por todo ello, para realizar los análisis posteriores, se descartaron los primeros dos espectros y los dos últimos.

La comparación de los espectros de una misma muestra adquiridos con cada una de las fibras situadas a la misma altura respecto al eje y (fibras 43,43b y 51) permitió determinar las similitudes o diferencias de la información obtenida a distintas posiciones respecto a la fuente de radiación (eje x). A modo de ejemplo, en la Figura 7.7 se muestran los espectros y el gráfico de *scores* al aplicar un PCA de una muestra de producción del producto granulado con un 5% de API. Con el fin de reducir las diferencias debidas a la variabilidad física de la muestra, se aplicó como pretratamiento el SNV. Los espectros obtenidos con las fibras que se encontraban a una misma distancia de la fuente de iluminación (en este caso 43 y 43b) eran equivalentes. Dado que los resultados que se obtengan de ellas serán comparables, estas pueden usarse para determinar la homogeneidad del producto que se esté analizando. Por otro lado, cuanto mayor es la distancia en el eje de las x entre la fuente de luz y el detector, en este caso fibra 51, la información proviene de capas más profundas de la muestra⁴. En los espectros de la fibra 51 se observaba una mayor presencia de ruido, sobre todo en el rango de 1470 a 1700nm cuando se compara con el espectro de la fibra 43 y a la vez, un efecto multiplicativo distinto, que podría estar causado por el distinto recorrido de la radiación en la muestra. Los *scores* también confirmaban estas diferencias. Cuantas más fibras a distintas distancias respecto a la fuente de luz se incluyan en el cálculo del modelo, mayor información de la muestra se obtendrá, debido a que el volumen analizado será mayor. En este caso, dado que solo había disponible la fibra 51, y no la 51b, el promedio de los espectros para desarrollar los modelos provendrían de las fibras 42, 43, 44, 42b, 43b y 44b.

Una característica remarcable de los espectros de todos los productos analizados, era la elevada presencia de ruido en el rango espectral 1425-1700 nm. Este rango no podría usarse en el desarrollo de los modelos cuantitativos debido a la aleatoriedad del mismo. Se confirmó que esta zona espectral es la que presentaba mayor variabilidad. Con el fin de evaluar la repetitividad del instrumento, se analizaron seis veces los mismos siete de comprimidos recubiertos de una muestra de producción. De cada repetición se calculó el espectro promedio para cada comprimido, siguiendo el esquema de la Figura 7.2. De los seis espectros obtenidos por comprimidos, se calculó el promedio y la desviación estándar asociada a cada longitud de onda. En la Figura 7.8 de la izquierda se representa esta desviación estándar para cada longitud de onda, poniendo de manifiesto que los valores de desviación estándar más elevados se obtenían en el rango espectral a partir de 1425nm. Este hecho se asoció a la presencia de ruido en esta zona. En la misma Figura 7.8 a la derecha aparece representado el espectro promedio de las seis repeticiones hechas del primer comprimido y su desviación estándar asociada. En el resto del espectro (por debajo de 1400nm) la variabilidad entre medidas era muy baja.

7.3.3. Respuesta del NIR a variaciones de concentración de API

El desarrollo de modelos cuantitativos con PLS permitió determinar si la respuesta del NIR evaluado era sensible a los cambios de concentración de cada uno de los productos escogidos. No se obtuvieron los mismos resultados para cada uno de ellos tal como se detalla a continuación.

El instrumento demostró ser sensible a la diferencia de concentración de API en dos productos: el granulado con un 5.0% de API y los comprimidos recubiertos con un 10.0% de API. Las figuras de

mérito de ambos modelos PLS se presentan en la Tabla 7.2. Los pretratamientos escogidos ordenaban las muestras según su concentración en la banda principal del API (1134nm en el producto granulado y 1140nm en los comprimidos). Además, en ambos casos el primer factor latente ordenaba las muestras según las concentraciones. Los resultados no eran satisfactorios para implementar los modelos en rutina pero permitieron confirmar la utilidad del instrumento para la determinación del contenido de API en ambos productos.

Tabla 7.2 Figuras de mérito de los modelos de calibración PLS.

Características	PLS cuantificación API 5.0% producto granulado con un	PLS cuantificación API 10.0% comprimidos recubiertos con un
Número de muestras (Laboratorio + Producción)	15 (12+3)	15 (12+3)
Pretratamiento espectral	Normalización vectorial y 2 ^a Derivada SG (polinomio orden 2, ventana móvil 13 puntos)	SNV y 2 ^a Derivada SG (polinomio orden 2, ventana móvil 15 puntos)
Rango espectral (nm)	1070-1180	1100-1490
Rango de calibración (mg/g)	33.8-65.8	66.2-140.7
Número de factores PLS	3	3
Varianza explicada (Y) (%)	99.1	96.2
RMSEC/RMSECV (mg/g)	1.03/1.80	4.68/6.79

La implementación del equipo en el proceso de dosificación, no solo permitiría la determinación del contenido de API, también sería posible determinar la homogeneidad de la muestra si se desarrollan modelos individuales para cada una de las fibras. Se puede calcular la desviación estándar entre los valores de concentración de API predichos con los modelos de fibras equidistantes respecto a la fuente de luz (p. ej. 43 y 43b) o bien entre los valores predichos para una misma fibra en función del tiempo. Sería necesario establecer un límite de control por debajo del cual se considera que el producto es homogéneo⁷. Con los datos obtenidos hasta el momento no ha sido posible realizar dicho estudio.

No fue posible desarrollar ningún modelo cuantitativo que fuera capaz de relacionar la señal con la concentración de API para el producto granulado con un 1.0% de API y el producto en polvo con un 5.6% de API. En el primer caso fue debido a la baja concentración de API en las muestras, del 0.7% al 1.3 % en peso. Aunque la banda del API a 1140nm no tiene interferencia con la señal del placebo, el instrumento, en las condiciones experimentales definidas, no era suficientemente sensible para llegar a diferenciar las muestras por la diferencia de concentraciones. Por otro lado, en el caso del producto en polvo que tiene un 5.6% en peso de API, su banda principal se encuentra en el inicio de la zona más ruidosa del espectro (1426nm), como se muestra en la Figura 7.9. No se encontró ninguna combinación de pretratamientos que permitiera reducir la aleatoriedad de la señal. La otra banda del API que aparece sin interferencias de la matriz a 970nm, no tenía

suficiente intensidad. Como resultado, no se pudo desarrollar ningún modelo PLS que demostrara que la respuesta NIR variaba según el contenido de API.

7.3.4. Detección de comprimidos rotos

El instrumento NIR multipunto también puede usarse a modo de cámara hiperespectral pero con datos simplificados. La comparativa del número de espectros registrados para un comprimido permitiría detectar muestras rotas. Si para un comprimido se conoce de antemano el número de espectros que debe registrar cada una de las fibras activadas -en función de la velocidad y del tiempo de integración- sería posible detectar comprimidos que están rotos, debido a que tendrían un número menor de espectros en algunos o todos los canales. A modo de ejemplo, en la Figura 7.10 de la izquierda se presenta el número de espectros obtenidos de unos comprimidos intactos, los cuales presentaban entre 24 y 27 espectros. En la figura de la derecha se muestran los espectros adquiridos para el mismo producto que anteriormente pero las 5 primeras muestras corresponden a comprimidos rotos. Se puede observar como el número de espectros registrados en la fibra 43 van de 15 a 23, confirmándose la disminución de espectros registrados por comprimido. En este caso, se podría instalar un sistema que expulsara con aire comprimido todos los comprimidos que tuvieran un número de espectros menor a 23 espectros.

7.4. CONCLUSIONES

Se han evaluado las capacidades que tiene un instrumento NIR con sonda multipunto para poder implementarse en el proceso de dosificación de distintos productos farmacéuticos. Solo se ha demostrado la validez de detectar cambios de concentración de API en dos, un producto granulado con un 5.0% de API y un comprimido recubierto que contiene un 10.0% de API. La sensibilidad del equipo, en las condiciones experimentales escogidas, no ha sido suficiente para detectar cambios en un producto con una baja concentración de API del 1.0%. El hecho que el intervalo espectral a partir de 1425nm presente un elevado ruido, ha imposibilitado cuantificar el API de un producto que tiene la banda más intensa a 1426nm. Este estudio de viabilidad necesita una segunda fase, donde se puedan optimizar las condiciones experimentales e incluir más fibras en las medidas, lo que podría suponer conseguir resultados más óptimos en los dos últimos casos. También sería necesario ampliar el estudio a otros parámetros de calidad como son el contenido en agua o la dureza de los comprimidos.

La mayor ventaja de los equipos multipunto es que se pueden usar como un equipo NIR tradicional, obteniendo el promedio de los espectros que corresponden a una unidad de dosis (por ejemplo a un comprimido) con el que se puede obtener información de parámetros de calidad como son la riqueza. Pero también permiten obtener información de la homogeneidad de la unidad de dosis al analizar de manera individual los espectros de cada fibra. Cabe mencionar que esta ventaja solo es beneficiosa en el caso de los comprimidos dado que en el análisis de productos granulados o polvo esta información de homogeneidad se puede obtener con los instrumentos NIR tradicionales, comparando los espectros obtenidos en función del tiempo. También son útiles para detectar la presencia de comprimidos rotos en línea. Una de las desventajas que se ha corroborado con este trabajo ha sido la dificultad del tratamiento de los datos. De momento, la exploración de datos y el desarrollo de los modelos lo debe realizar una persona experta en quimiometría y con conocimiento moderado del uso de un software de procesamiento de datos (como son el Matlab o R). Procedimentar dicho método y comunicar a las autoridades sanitarias todos el desarrollo es un reto.

7.5. BIBLIOGRAFÍA

1. European Pharmacopoeia. Ph. Eur. 9.0 Chapter 2.9.47. Demonstration of uniformity of dosage units using large sample sizes. (2017).
2. Burdick, R. K. *et al. Statistical Applications for Chemistry, Manufacturing and Controls (CMC) in the Pharmaceutical Industry.* (Springer, 2017).
3. Igne, B., Talwar, S., Feng, H., Drennen, J. K. & Anderson, C. A. Near-Infrared Spatially Resolved Spectroscopy for Tablet Quality Determination. *J. Pharm. Sci.* **104**, 4074–4081 (2015).
4. Boiret, M. & Chauchard, F. Use of near-infrared spectroscopy and multipoint measurements for quality control of pharmaceutical drug products. *Anal. Bioanal. Chem.* **409**, 683–691 (2017).

Estudio de viabilidad de un método NIR para implementar en la etapa de dosificación

5. Burns, D. A. & Ciurczak, E. W. *Handbook of Near Infrared Analysis*. (CRC Press Taylor and Francis, 2008).
6. Siesler, H. W., Ozaki, Y., Kawata, S. & Heise, H. M. *Near-infrared spectroscopy. Principles, instruments, applications*. (Wiley-VCH, 2002).
7. Shi, Z., Cogdill, R. P., Short, S. M. & Anderson, C. A. Process characterization of powder blending by near-infrared spectroscopy: Blend end-points and beyond. *J. Pharm. Biomed. Anal.* **47**, 738–745 (2008).

Parte IV

VALIDACIÓN Y VERIFICACIÓN DE LIMPIEZA

8. IMPLEMENTACIÓN DE LÍMITES PARA MÉTODOS NO ESPECÍFICOS

8.1. INTRODUCCIÓN

La limpieza de los equipos, instalaciones y locales es una etapa crítica dentro de la industria farmacéutica que debe prevenir la contaminación microbiológica y la contaminación cruzada^{1,2}. Cualquier procedimiento de limpieza debe ser validado, para aportar las evidencias de que el proceso establecido es adecuado para eliminar cualquier residuo que haya estado en contacto con el equipo, instalación o sala³. Cuando la planta de producción es multiproducto, este es un reto aún mayor. Una vez el proceso de limpieza está validado, puede ser necesaria la verificación de dicha limpieza para asegurar que se está realizando correctamente. En los medicamentos que formulan sustancias activas de elevada y/o especial actividad, la verificación de la efectividad de la limpieza es obligatoria.

Un producto puede ser contaminado por el producto que se ha fabricado previamente en el equipo (tanto API como excipientes), por agentes microbiológicos, por contaminantes físicos (como polvo o partículas), por lubricantes, por los agentes de limpieza utilizados (detergentes y/o desinfectantes) o por residuos de descomposición⁴. Es necesario demostrar con la validación de limpieza de los equipos, instalaciones o locales que, una vez finalizado el proceso, todos los contaminantes están por debajo de los niveles aceptables. Estos límites se establecen basándose en una evaluación de la toxicidad³. El primer criterio de cálculo es el que se realiza a partir de la exposición diaria permitida (PDE por sus siglas en inglés *Permitted Daily Exposure*)⁵. Si no se dispone de la información del PDE, se debe calcular basándose en dosis terapéuticas o en el valor de dosis letal media (LD₅₀). En última instancia, si el contaminante no es tóxico, cada laboratorio debe definir y justificar el límite aceptable. Previamente a la aplicación de los valores de PDE ha sido una práctica habitual y aceptada por las agencias reguladoras aplicar como límite genérico del potencial contaminante en 10ppm. El límite de microorganismos aceptables se establece según la finalidad del medicamento.

Para llevar a cabo la validación de limpieza, deben determinarse tanto el método de muestreo (por hisopo o en el agua purificada del último aclarado) como la técnica analítica a utilizar para detectar los residuos definidos. Siempre que sea posible es necesario establecer un método analítico específico para cada contaminante. Las técnicas analíticas específicas más empleadas, las que presentan señal específica para un determinado analito, son la cromatografía HPLC, la cromatografía de gases (GC) y la espectroscopia UV o IR. Por otro lado, cuando no existe una técnica analítica específica, se acepta el uso de técnicas no específicas que determinan todos los

analitos que presentan señal como son las que determinan el carbono orgánico total (TOC por sus siglas en inglés *Total Organic Carbon*), el pH o la conductividad. Las ventajas de estas últimas son la rapidez de implementación, la instrumentación es relativamente barata y se pueden instalar de modo on-line.

Se consideran agentes de limpieza el agua: productos químicos simples (ácidos o bases diluidas), detergentes, agentes antimicrobianos o disolventes orgánicos. No es fácil conocer la composición de los detergentes usados durante el proceso de limpieza de equipos debido a la poca información que proporcionan los proveedores, lo que hace complicada la tarea de evaluar los residuos⁶. Por lo que es habitual usar técnicas no específicas para determinar la presencia de estos contaminantes. Cuando se usan métodos de análisis no específicos, se está sujeto a tener una respuesta de todo aquello que sea sensible a la técnica. No se puede asegurar, a priori, cual es el analito o analitos que presentan señal. Para ello, es necesario hacer un estudio de todos los posibles contaminantes presentes en una muestra que puedan dar respuesta. Para disminuir el número de experimentos y obtener la información de la máxima calidad se puede usar un diseño de experimentos.

El objetivo de este trabajo ha sido establecer una metodología que permita entender el comportamiento de todos los componentes relevantes involucrados en los procesos de limpieza de los distintos equipos usados para la fabricación de un medicamento, frente a una técnica analítica no específica. Esto permitiría crear un perfil del proceso de eliminación de agente de limpieza y así poder seguir de manera *on-line* o *at-line*, dependiendo del método de muestreo, la limpieza de los equipos y verificar la misma de manera rápida, fácil y económica. Para ello, se han evaluado todos los agentes de limpieza usados en las instalaciones de producción. También se ha elegido un medicamento determinado y se han seguido todos los procedimientos de limpieza de los equipos implicados en su producción, con el objetivo de proponer una metódica para estudiar el comportamiento de los factores a la respuesta analítica. Finalmente, para un segundo medicamento, se ha determinado la respuesta máxima permitida en cada uno de los equipos una vez finalizada la limpieza. La técnica analítica escogida ha sido la conductividad. La conductividad es una técnica no específica que permite establecer la cantidad de especies iónicas presentes en una muestra, ya que mide la facilidad de los iones al conducir la electricidad. La conductividad de una disolución depende de la concentración de iones, su movilidad y carga, la temperatura y viscosidad de la disolución.

8.2. MÉTODO EXPERIMENTAL

8.2.1. Materiales y equipo

En total se usaron diez agentes de limpieza. Los detergentes Mistol Vajillas (Henkel Ibérica; España), Skip Polvo y Skip Líquido (Unilever, Reino Unido), CIF profesional crema (Diversey,

España), Divoflow NTC y Shureclean Plus (JohnsonDiversey, España); el desinfectante Divosan Mezzo y Sumasol (Diversey, España); el desincrustante Pascal VA5 (JohnsonDiversey, España); y el antiespumante AF silicon (JohnsonDiversey, España).

El primer medicamento, que sirvió para establecer el mecanismo de trabajo para entender el comportamiento de los distintos factores frente a la conductividad, fue un comprimido recubierto que contenía un 13% en peso de API y 6 excipientes entre los que se encontraban celulosa microcristalina, almidón de maíz y carboximetilalmidón sódico como mayoritarios. El segundo medicamento estudiado, el cual sirvió a modo de ejemplo de cómo determinar las conductividades máximas permitidas una vez finalizados todos los procesos de limpieza, fue un granulado para solución oral que contenía un 5% de API y 6 excipientes, de los cuales la sacarosa y el ácido cítrico son los más relevantes.

Para la preparación de las disoluciones pertinentes se empleó metanol de calidad HPLC $\geq 99.9\%$ (Carlo Erba Reagents, España) y agua purificada de la red de distribución de la planta de producción farmacéutica. Esta última también se utilizó para medir los blancos de la conductividad.

La conductividad se midió con el conductímetro modelo 712 (Metrohm, Suiza).

8.2.2. Preparación y medida de muestras

Para la evaluación inicial de la respuesta de cada uno de los 10 agentes de limpieza, se prepararon disoluciones en un rango de 1 a 100 ppm con agua purificada. De cada muestra, se usaron 25mL de solución para medir las conductividades.

Las muestras preparadas para realizar el DoE relacionado con el primer medicamento, que sirvió para determinar los factores que tenían efecto en la conductividad, podían contener API, excipientes, detergente e hisopo (*swab*). Las disoluciones se prepararon de acuerdo a las concentraciones indicadas por el DoE calculado en cada caso, usando como solvente el agua purificada. Si era pertinente, se mantenía la solución final en ultrasonidos durante un minuto. Se usaron 25mL para determinar la conductividad de cada muestra.

El estudio de recuperación de detergente, en este caso concreto de Mistol, para cada superficie de contacto al muestrearse con *swab*, se realizó como sigue y como se esquematiza en la Figura 8.1. Se preparó una disolución madre de 4000ppm de Mistol en metanol. Para las muestras de recuperación, se depositaron 200, 250 y 300 μL en un área de 25cm^2 en cada una de las superficies estudiadas (acero inoxidable mate y acero inoxidable brillante). Al mismo tiempo, se transfirieron estas cantidades en tres matraces distintos, los cuales servirían como muestras de control. Una vez evaporado el solvente de las superficies, se muestrearon los 25cm^2 de cada una de ellas con un bulbo empapado en agua purificada, el cual se cortó y se introdujo en un matraz. En todos los

matraces, también en los de control, se añadieron 25mL de agua purificada y se agitaron vigorosamente durante un minuto. Las concentraciones de las disoluciones finales fueron 32, 40 y 48 ppm (80, 100 y 120 % respectivamente en relación a la máxima cantidad permitida en la superficie de uno de los equipos estudiados). Acto seguido, se introdujeron durante un minuto en el baño de ultrasonidos y se midió la conductividad⁷ tanto de las disoluciones provenientes del

muestreo como de las disoluciones iniciales, denominadas de control.

Las muestras que permitieron establecer la conductividad límite esperada en los equipos de producción del segundo medicamento se prepararon con los contaminantes que tenían efecto en la conductividad. La concentración de cada uno de ellos dependía de los siguientes supuestos: la recuperación del material del equipo, el nivel aceptable de contaminante en la superficie del equipo o en las aguas del último aclarado (según el muestreo especificado) y, en el caso de muestreo con *swab*, del factor de dilución. La conductividad se midió con 25mL de cada disolución.

8.2.3. Definición de los factores, de la respuesta y niveles para el DoE

Dado que el objetivo del DoE era realizar un *screening*, se hizo un estudio de todos los materiales y agentes de limpieza usados en cada uno de los equipos involucrados en la producción de los

medicamentos. Si la limpieza de cada equipo se realizaba con detergentes distintos, se definía un DoE para cada uno de ellos. Los resultados del DoE permitieron determinar qué factores tenían efecto en la conductividad. En el caso de que más de un equipo se limpiase con el mismo producto, entonces se escogió el que estaba en contacto con un mayor número de componentes de la formulación.

Se identificaron como factores los componentes del medicamento (API y excipientes), el detergente y el tiempo de exposición al aire de la solución a analizar. Los niveles incluidos en el estudio fueron la ausencia de cada factor (-) y el máximo esperado de cada factor (+). La respuesta del estudio era la conductividad que presentaba la disolución final.

El cálculo y análisis del DoE se realizó con el software Modde 8.0 (Umetrics, USA). Una vez finalizada la etapa experimental, se analizaron los datos obtenidos para poder determinar los factores que tenían efecto en la conductividad. Los modelos para relacionar la conductividad con los factores significativos se calcularon usando el algoritmo de regresión lineal múltiple.

8.3. RESULTADOS Y DISCUSIÓN

8.3.1. Evaluación inicial de los agentes de limpieza

En la Tabla 8.1 se presentan los rangos de concentración de cada producto de limpieza evaluados en este estudio, con los cuales se determinó la idoneidad del uso de la conductividad para detectar su presencia en una solución acuosa. El intervalo de linealidad se estableció visualmente, ya que el objetivo de este estudio previo era obtener una idea inicial del comportamiento de cada uno de los 10 productos de limpieza más usados en la planta de producción. El límite de detección se calculó a partir de dividir tres veces la desviación estándar de la intercepción de la recta de regresión y la pendiente calculada.

Tabla 8.1 Rangos estudiados, intervalo de linealidad y LOD de cada producto de limpieza.

Nombre comercial	Rango estudiado (ppm)	Intervalo de linealidad (ppm)	LOD (ppm)
Mistol Vajillas	1.9-111.8	5.7-111.8	4.2
Jabón Skip polvo	1.0-99.4	1.0-99.4	0.8
Jabón Skip líquido	2.3-111.1	22.2-111.1	6.1
Divosan Mezzo	1.3-111.1	3.3-111.1	1.9
CIF profesional crema	1.1-99.3	5.6-99.3	3.8
Pascal VA5	1.3-32.7*	1.3-32.7	0.4
Divoflow NTC NTC	1.1-105.3	8.3-105.3	1.5
AF silicon	0.8-102.0	-	-

Tabla 8.1 Rangos estudiados, intervalo de linealidad y LOD de cada producto de limpieza.

Nombre comercial	Rango estudiado (ppm)	Intervalo de linealidad (ppm)	LOD (ppm)
Sumasol	1.1-101.0	1.1-101.0	0.3
Shureclean plus	1.4-108.8	10.2-108.7	6.8

*La concentración máxima estudiada es 32.7ppm porque presenta la conductividad límite del instrumento, 100 μ S/cm.

Los resultados obtenidos confirmaron la idoneidad de la técnica analítica para poder detectar, a priori, la presencia de todos los agentes de limpieza a excepción del AF silicon. En este caso, debería usarse un método alternativo. Los tensioactivos basados en silicona están formados por dos partes. Un pequeño grupo polar y una cadena polimérica hidrofóbica que suele ser polidimetilsiloxano⁸. Si este fuera uno de los ingredientes del producto AF silicon, el TOC sería posiblemente un método analítico válido para su determinación. La reciente adquisición, por parte de la empresa, de un equipo de determinación del TOC permitirá comprobar dicha hipótesis.

8.3.2. Estudio de la influencia de factores en la conductividad

En este apartado se define la metodología a seguir para identificar todos aquellos factores, tanto los que están implicados en el proceso de limpieza de los equipos usados en la producción de un medicamento como los que están relacionados con su análisis, que tienen efecto en la conductividad. Esta debe permitir justificar el uso de la técnica no específica para el propósito de determinar la presencia de detergente y reducir el número de componentes que se deberán usar en dos posibles estudios posteriores. Uno de ellos sería describir el perfil del proceso de limpieza mediante conductividad y el otro fijar la conductividad límite que pueden tener las disoluciones finales y así verificar la correcta finalización de la limpieza. A modo de ejemplo se muestra el estudio realizado en el proceso de producción del primer medicamento, los comprimidos recubiertos, y sus respectivos procesos de limpieza.

8.3.2.1. Identificación de factores

Cada equipo, o las distintas partes desmontables del mismo, que interviene en la producción de un medicamento tienen su propio proceso de limpieza. En estos procedimientos, entre otros, se definen los modos de limpieza (automático o manual) y los agentes de limpieza. El primer paso que se hizo para evaluar la idoneidad de la conductividad con el fin de verificar las limpiezas fue recopilar y esquematizar todo lo descriptivo del proceso. En la Figura 8.2 se detallan las 5 operaciones unitarias que conforman la producción de los comprimidos recubiertos, los equipos implicados, los productos de la formulación que están en contacto con las superficies de estos equipos y finalmente, los detergentes usados en los procesos de limpieza de cada parte. Se planificó un DoE para cada producto de limpieza usado. En el caso de que más de un equipo se limpiase con el

mismo detergente, se escogió el que estaba en contacto con más componentes de la formulación. Los 3 equipos que se estudiaron están resaltados en naranja en la Figura 8.2.

Los factores definidos en cada caso fueron: los componentes que estaban en contacto con cada equipo (API y excipientes), el detergente, el *swab* y el tiempo de exposición de la solución final al aire. Este último se incluyó debido a que puede ser determinante el tiempo transcurrido entre la preparación u obtención de la muestra y el análisis. El dióxido de carbono ambiental se disuelve en el agua purificada de las muestras alterando la conductividad de las mismas. La concentración de bicarbonato presente en el agua depende del tiempo de contacto agua-aire.

8.3.2.2. Determinación de niveles

Los niveles inferiores de cada factor se definieron como la ausencia de cada uno de ellos en la disolución. En cambio, los niveles superiores de cada factor se establecieron de manera distinta según sus características, tal como se explica a continuación.

Para el API se disponía del valor PDE, por lo que la cantidad máxima de contaminante que puede pasar al lote siguiente (MACO por sus siglas en inglés *Maximum allowable carryover*) en cada equipo se podía calcular según la ecuación Ec 8.1. Disponiendo del valor MACO y la superficie del equipo que está en contacto con el producto, se calcularon los niveles máximos de API permitidos en la superficie de cada equipo una vez finalizada la limpieza.

$$\text{Nivel máximo API superficie equipo} = \frac{MACO_{PDE}}{ES} = \frac{PDE \cdot MBS}{TDD \cdot ES} \quad \text{Ec 8.1}$$

donde las unidades de PDE se expresan en mg/día, el tamaño de lote mínimo del siguiente producto (MBS) en mg, la dosis máxima diaria del siguiente producto (TDD) en mg/día y la superficie del equipo (ES) en cm².

Para los excipientes no se disponía de información toxicológica. Es por ello que se usó el límite interno, el cual establece que en el siguiente producto no puede haber más de 1ppm de excipiente. Este límite es una restricción del criterio empleado históricamente de 10ppm para los productos sin efecto tóxico el cual permite disponer de un factor de seguridad. El nivel máximo en la superficie del equipo se determinó siguiendo la ecuación Ec 8.2.

$$\text{Nivel máximo Excipientes superficie equipo} = \frac{1\text{ppm} \cdot \text{MBS}}{\text{ES}} \quad \text{Ec 8.2}$$

donde MBS está expresado en mg y ES en cm^2 .

En el caso de los detergentes no existe ninguna especificación de la EMA o la FDA que indiquen el nivel máximo permitido al finalizar el proceso. Se afirma que idealmente no debería estar presente al acabar el ciclo de limpieza, pero dejan a criterio de cada compañía establecer el límite⁶. Una práctica habitual es utilizar el criterio de un máximo de 10ppm en el producto siguiente⁷. En cada caso, si en la ficha de seguridad del producto aparecía información del LD_{50} , primero se calculó el $\text{MACO}_{\text{LD}_{50}}$ según la ecuación Ec 8.3. En ninguno de los cuatro detergentes se proporcionaba la información toxicológica del producto global. En las fichas de seguridad de cada uno, se indicaba la composición máxima que podía haber de aquellos ingredientes que presentaban LD_{50} y el valor de este parámetro. Con estos dos datos se calculó una LD_{50} aparente del producto global, el cual se empleó para calcular el $\text{MACO}_{\text{LD}_{50}}$.

$$\text{MACO}_{\text{LD}_{50}} = \frac{\text{NOEL} \cdot \text{MBS}}{\text{SF} \cdot \text{TDD}} = \frac{\text{LD}_{50} \cdot \text{BW} \cdot \text{MBS}}{2000 \cdot \text{SF} \cdot \text{TDD}} \quad \text{Ec 8.3}$$

donde el LD_{50} se expresa en mg/kg, el peso de una persona adulta estándar (BW) es de 70kg, MBS en mg, 2000 es una constante sin unidades, el factor de seguridad (SF) es de 100 para medicamentos orales y es adimensional, TDD en mg/día y ES en cm^2 .

El $\text{MACO}_{\text{LD}_{50}}$ se comparó con el $\text{MACO}_{10\text{ppm}}$, calculado a partir de la ecuación Ec 8.4. Entre los dos valores, se escogió el MACO más restrictivo para calcular el nivel de detergente que podía quedar en la superficie del equipo una vez finalizara el ciclo de limpieza (ecuación Ec 8.5).

$$\text{MACO}_{10\text{ppm}} = 10\text{ppm} \cdot \text{MBS} \quad \text{Ec 8.4}$$

donde MBS se expresa en mg.

$$\text{Nivel máximo Detergente superficie equipo} = \frac{\{\min\{\text{MACO}_{\text{LD}_{50}}|\text{MACO}_{10\text{ppm}}\}\}}{\text{ES}} \quad \text{Ec 8.5}$$

donde ES se expresa en cm^2 .

En todos los casos anteriores, a partir de los niveles máximos establecidos, se calcularon las concentraciones que tendrían las soluciones finales suponiendo que las superficies de los equipos se muestreaban por *swab*. Así, cada valor del nivel máximo permitido en la superficie de los equipos se multiplicó por el área muestreada (25cm²) y se dividió por el volumen de agua purificada usado para la extracción del *swab*, 25mL. En la Tabla 8.2 se detallan las concentraciones finales. Para la línea de blísteres se consideró que debido a la baja concentración de los excipientes 5 y 6, estos podían descartarse del estudio. Previamente se había comprobado que estas sustancias no presentaban conductividad.

Tabla 8.2 Concentraciones máximas que pueden quedar en la solución final una vez finalizado el ciclo de limpieza cuando la superficie del equipo se ha muestreado con *swab*.

Factores	Granuladora (ppm)	Comprimidora parte A (ppm)	Comprimidora parte B (ppm)	Línea blísteres (ppm)
API	3.75	4.00	4.00	4.00
Excipiente 1	14.57	15.53	15.53	15.53
Excipiente 2	21.40	22.82	22.82	22.82
Excipiente 3	15.71	16.75	16.75	16.75
Excipiente 4	NA	2.36	2.36	2.36
Excipiente 5	NA	NA	NA	0.03
Excipiente 6	NA	NA	NA	0.00
Skip polvo	6.64	NA	NA	NA
Divoflow NTC	NA	NA	17.36	NA
Skip líquido	NA	6.44	NA	NA
Mistol	NA	NA	NA	4.14

En el caso del factor *swab*, que es una variable cualitativa dicotómica, el nivel máximo era la presencia de él en la solución final seguido de aplicarle ultrasonidos durante un minuto.

Finalmente, para el factor tiempo de exposición de la solución al aire, el nivel máximo definido fue de una hora de transcurso entre la preparación de la solución y la lectura por conductividad.

8.3.2.3. Elección del DoE

El número de factores definidos en la granuladora fueron 7, mientras que para la comprimidora parte A y B, y la línea de blísteres fueron 8. Los DoE calculados para cada equipo, o parte de ellos, fueron diseños factoriales fraccionados de 2 niveles, con una resolución IV y tres puntos centrales. Estos permitían obtener información clara de los efectos principales sin tener una carga muy elevada de experimentos; consiguiendo un buen compromiso para realizar un *screening*. En total se planificaron 19 experimentos para cada equipo de producción. Las preparaciones de muestra y

lecturas de la conductividad se llevaron a cabo de forma aleatorizada para evitar posibles influencias sistemáticas.

8.3.2.4. Análisis de los resultados del DoE

El modelo de los resultados obtenidos para el DoE de la granuladora se calculó con 18 experimentos de los 19 propuestos. Debido a un error experimental, un punto central se tuvo que excluir del modelo debido a que la conductividad obtenida fue de -1.13 y el valor se consideró inválido. Se podía confiar con los resultados obtenidos porque el modelo presentaba un buen ajuste, con una varianza explicada (R^2) de 0.999 y un poder de predicción (Q^2) de 0.996; y la regresión era significativa (0.000) mientras que la falta de ajuste era no significativo (0.74). Los coeficientes significativos del modelo (con un valor de $p < 0.05$) fueron los correspondientes al API, al detergente Skip polvo y al tiempo de exposición de la solución al aire; por lo que estos se identificaron como los factores que tenían efecto en la conductividad. En la Figura 8.3 se presentan gráficamente los coeficientes. Aquellos en los que la línea de incertidumbre cruza el 0 del eje de la y, indica que el factor asociado no tiene influencia en la respuesta. Cabe mencionar que la eliminación de los factores con menor significación (excipiente 1 y 2 y el *swab*) no hizo que el factor Excipiente 3 pasara a ser significativo.

El modelo obtenido para la parte A de la comprimidora también presentaba un buen ajuste; con una R^2 de 0.968 y una Q^2 de 0.901, la regresión era significativa (0.000) y la falta de ajuste no (0.06). El estudio puso de manifiesto que los factores con efecto en la conductividad eran el excipiente 3 y el detergente Divoflow NTC (Figura 8.5). El excipiente 3 correspondía al carboximetilalmidón sódico.

En el caso de la parte B de la comprimidora, un modelo con una R^2 de 0.977 y una Q^2 de 0.920 con la regresión significativa (0.00) y la falta de ajuste no significativo (0.43) desveló que los únicos factores que presentaban un coeficiente significativamente distinto a 0 (valor de $p < 0.05$) eran el API, el excipiente 3 y el tiempo de exposición. Los coeficientes calculados se encuentran en la Figura 8.4. En este caso concreto, el detergente Skip líquido no era un factor con efecto en la conductividad. El nivel máximo permitido, 6.44ppm (Tabla 8.2), está muy cerca del límite de detección calculado en el estudio previo (Tabla 8.1) para este detergente, el cual era 6.1ppm. Para este proceso de limpieza la conductividad no es una técnica adecuada para determinar la presencia de detergente por debajo del nivel de aceptación. Se propuso como técnica analítica alternativa el TOC.

Finalmente, en la línea de blísteres, donde el detergente usado era el Mistol, confirmó que este era un factor con efecto a la conductividad, además del API y el excipiente 3 (Figura 8.6). El tiempo no fue significativo una vez eliminados los factores con un coeficiente menor (Excipiente 1, 2 y 4 y el

swab). El modelo presentaba una R^2 de 0.970 y una Q^2 de 0.884, siendo la regresión significativa (0.00) y la falta de ajuste no (0.70).

Este estudio permitió establecer los factores a incluir en la preparación de muestras que servirían para establecer el perfil del proceso de limpieza o determinar la conductividad límite para verificar que dicho proceso había finalizado de manera correcta. También sirvió para poner de manifiesto que no era posible usar la técnica de la conductividad para determinar todos los detergentes. Fue necesario proponer un método alternativo para el análisis del Skip líquido.

8.3.3. Establecimiento conductividad límite para verificar la limpieza

En este apartado se describe todo el estudio realizado para establecer la conductividad límite que tendrían las disoluciones finales de las muestras recogidas al finalizar el proceso de limpieza. Estos límites servirían para verificar los procesos de limpieza respecto al contaminante detergente. El ejemplo que se presenta es el realizado con el segundo medicamento, un producto granulado.

8.3.3.1. Factores que influyen en la conductividad

Para iniciar el estudio, el primer paso fue determinar todos los posibles factores que podían afectar a la conductividad de las soluciones finales. Para ello, se empleó la metodología establecida en el apartado 1.3.3. Todos los equipos usados en la producción del segundo medicamento se limpian con el mismo detergentes tal como de muestra en la Figura 8.7. En este caso, fue suficiente realizar un solo DoE para obtener la información buscada. El equipo seleccionado fue la línea de sobres, puesto que era el que estaba en contacto con todos los componentes de la formulación.

Se definieron como factores a evaluar todos los componentes de la formulación (API y excipientes), el detergente Mistol y el tiempo de contacto de la solución con el aire. El factor *swab* se descartó, debido que en el estudio anterior ya se había comprobado que, tanto el mismo hisopo como el hecho de aplicar ultrasonidos, no tenía influencia en la conductividad. En total se definieron 9 factores, con los niveles inferiores y superiores calculados de la misma manera que en el apartado 1.3.2.2.

El DoE calculado fue un diseño factorial fraccionado de resolución IV, con un total de 33 experimentos planificados. Tras la evaluación y análisis de los resultados experimentales se obtuvo un modelo con una R^2 de 0.959 y una Q^2 de 0.914 como parámetros de ajuste. Los coeficientes significativos del modelo (valor $p < 0.05$) fueron los que correspondían al excipiente 4, el Mistol y el tiempo. El excipiente 4 era el ácido cítrico. En la Figura 8.8 se muestra la representación gráfica de cada coeficiente. La eliminación de los factores no significativos (excipiente 1, 2, 3, 5 y 6) no provocó que el API se convirtiera en un factor significativo. Los tres factores comentados inicialmente fueron los que se tuvieron en cuenta para determinar la conductividad límite.

8.3.3.2. Determinación de la recuperación en superficies

Otro parámetro a determinar, antes de poder definir las conductividades límites al finalizar la limpieza de los equipos, fue el factor de recuperación del detergente en las distintas superficies de los equipos. El modo de muestreo estudiado en este caso fue el *swab*, por lo que el proceso de

recuperación se evaluó como el conjunto del muestreo, la extracción del *swab* y el método analítico.

Las superficies de los equipos del proceso de producción del segundo medicamento son acero inoxidable mate y brillante. Para cada superficie y nivel de concentración de detergente se procedió por triplicado. El porcentaje de recuperación se calculó como el cociente entre la respuesta de la conductividad de las muestras de recuperación y la solución de control. En la Tabla 8.3 se muestran los promedios y la desviación estándar relativa (RSD por sus siglas en inglés *Relative Standard Deviation*). En todos los casos se obtuvo una recuperación mayor al 50%, la cual se considera aceptable^{6,9}. Para asignar la recuperación del Mistol en cada superficie, se escogió el peor caso de cada una de ellas y se redondeó por debajo. Al acero inoxidable mate se le asignó una recuperación del 80% mientras que para el brillante fue del 50%.

Tabla 8.3 Resultados del proceso de recuperación de detergente Mistol en dos superficies, acero inoxidable mate y acero inoxidable brillante.

Superficie	Nivel de concentración detergente (%)	Promedio recuperación (%)	RSD (%)
Acero mate	80%	97.7	6.0
	100%	88.7	1.3
	120%	96.6	5.0
Acero brillante	80%	82.9	0.9
	100%	68.9	8.7
	120%	63.8	5.0

8.3.3.3. Conductividades límites

En la Tabla 8.4 se detallan las conductividades límite que podrán tener las soluciones finales de los equipos muestreados por *swab*. Por debajo de esta respuesta se podrá dar como correcta la limpieza del equipo respecto a la eliminación del detergente. Para obtener dichos valores se prepararon soluciones que contenían las concentraciones máximas de cada producto con efecto en la conductividad, teniendo en cuenta el nivel aceptable para cada uno de ellos según el MACO y el factor de recuperación del detergente en cada una de las superficies. Antes de proceder a la lectura, cada una de las soluciones se tapó y dejó reposar 15 minutos. Fue el tiempo máximo estimado que podría tardar un analista entre realizar la extracción y proceder a la lectura. Las conductividades descritas son las que se obtuvieron una vez restada la señal del blanco.

Tabla 8.4 Determinación de las conductividades límites en cada equipo usado en la producción de un producto farmacéutico granulado

Equipo	Tipo de muestreo	Superficie	Factor recuperación Misto (%)	Excipiente 4 (ppm)	Mistol (ppm)	Conductividad máxima (μS/cm)
Reactor	Último aclarado	Acero inoxidable Mate	--	--	--	--
Granulador	Swab	Acero inoxidable Brillante	50	0.90	49.98	6.576
Línea de sobres	Swab	Acero inoxidable Brillante	50	0.33	18.11	2.522
		Acero inoxidable Mate	80	0.33	28.98	3.555

En el caso del reactor, aún no se había realizado el estudio de recuperación del detergente con las aguas del último aclarado. La determinación de este parámetro es clave para poder establecer la conductividad máxima aceptable de la solución final.

8.4. CONCLUSIONES

Se desarrolló una metodología útil para establecer qué factores presentes en la disolución acuosa tras la limpieza de un equipo influyen en la conductividad. Realizar un DoE ha demostrado ser imprescindible cuando se trabaja con la conductividad como técnica analítica por varias razones. Entre ellas, reducir la carga experimental de futuros estudios, entender, como se ha comentado, qué factores influyen en la respuesta y asegurar que el analito objetivo a la concentración debida puede ser determinado con la técnica analítica. Esta metodología no es exclusiva para la conductividad ya que puede ser adecuada a cualquier otro método analítico no específico.

Una vez establecidas las variables que tienen efecto en la señal de la técnica analítica no específica, estas pueden usarse para establecer un perfil de proceso de limpieza. También es posible determinar la respuesta límite que podrían tener las muestras, recogidas por *swab* o las aguas del último aclarado, para dar como correcta la finalización de la limpieza respecto a la presencia de detergente. Esta metodología se puede realizar de modo *off-line* en el laboratorio de control de calidad o implementarla como un sistema PAT y verificar en tiempo real el proceso de limpieza.

Finalmente, se ha evidenciado que la conductividad no es adecuada para determinar la presencia de todos los agentes de limpieza presentes en la planta de producción. Puesto que la parte hidrofóbica de los tensioactivos suelen contener cadenas alifáticas o cualquier otra cadena con contenido de átomos de carbono, una técnica más adecuada podría ser el TOC.

8.5. BIBLIOGRAFÍA

1. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Capítulo 5 Producción. 11 (2015).
2. AEMPS. Guía de NCF de Medicamentos de Uso Humano y Veterinario, Capítulo 3: Locales y Equipos Bases. 1–5 (2015).
3. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Anexo 15 Cualificación y validación. (2015).
4. World Health Organization (WHO). Annex 4 Supplementary guidelines on good manufacturing practices : validation. 107–178 (2006).
5. European Medicines Agency. Guideline on setting health based exposure limits for use in risk identification in the manufacture of different medicinal products in shared facilities Draft. 11 (2014).
6. Food and Drug Administration (FDA). Guide to inspections validation of cleaning processes. (2010).
7. Milenović, D. M., Pešić, D. S. & Mitić, S. S. NON-SPECIFIC METHODS FOR DETECTING RESIDUES OF CLEANING AGENTS DURING. *Chem. Ind. Chem. Eng. Q.* **17**, 39–44 (2011).
8. Kekevi, B., Berber, H. & Yıldırım, H. Synthesis and Characterization of Silicone-Based Surfactants as Anti-Foaming Agents. *J. Surfactants Deterg.* **15**, 73–81 (2012).
9. Kaiser, H. J. & Ritts, B. Validation of Analytical Methods Used in Cleaning Validation. *J. Valid. Technol.* **10**, 15–30 (2004).

9. CARACTERIZACIÓN DE ALCOHOLES GRASOS LINEALES ETOXILADOS POR ESPECTROMETRÍA DE MOVILIDAD IÓNICA (IMS)

9.1. INTRODUCCIÓN

Ya se ha demostrado la habilidad de la espectrometría de movilidad iónica (IMS) como técnica específica a usar en las validaciones de limpieza para determinar la presencia de API y detergente de una manera rápida¹⁻⁴. Pero no se ha estudiado en profundidad el comportamiento del detergente frente a esta técnica. En este trabajo se estudian alcoholes lineales etoxilados (LAE del inglés *Linear Alcohol Ethoxylate*) los cuales son sustancias tensioactivas no iónicas. Los LAE tienen una gran importancia en la industria por sus propiedades tensioactivas, pues estos se usan en la preparación de detergentes, emulsionantes, humectantes, dispersantes, etc. Estos productos presentan dos grupos funcionales: la cadena hidrocarbonada (lipófila) de longitud diversa según el alcohol de partida y los grupos etoxi (hidrófila) polares, pero con polaridad variable según el número de grupos presentes en la molécula. La mayoría de los productos comerciales se presentan en forma de mezclas complejas de una o más cadenas hidrocarbonadas y con diferente número de grupos etoxi. La síntesis industrial -que consiste en hacer reaccionar el alcohol con óxido de etileno en la proporción deseada en presencia de un catalizador adecuado- da lugar a mezclas de oligómeros con diferentes valores promedio de grupos etoxi según la cantidad de óxido de etileno introducido en el reactor durante su fabricación. La importancia de conocer la composición de las mezclas obtenidas se debe a que sus propiedades dependen de la distribución obtenida y por lo tanto es deseable disponer de un método de análisis simple, rápido y reproducible para realizar la caracterización del producto obtenido con el objeto de realizar la selección del más adecuado para cada aplicación particular. Tradicionalmente, la caracterización de alcoholes etoxilados se ha llevado a cabo mediante técnicas de separación como cromatografía de gases (GC), cromatografía líquida de alta resolución (HPLC), cromatografía en capa fina (TLC) y electroforesis capilar (CE), las cuales requieren de patrones puros para la asignación de los picos. Además, en la mayoría de casos es necesario derivatizar la mezcla con reactivos adecuados, lo que da lugar a métodos largos y complejos que consumen mucho tiempo⁵⁻⁸. Usualmente estas técnicas se aplican acopladas a la espectrometría de masas (MS) para la caracterización de los diferentes oligómeros⁹⁻¹¹.

La separación en IMS se basa en la diferencia de velocidad de las especies cargadas bajo la acción de un campo eléctrico, que a su vez depende de su relación masa-carga y/o su estructura. En la ecuación Ec 9.1, la velocidad de deriva (v_d , cm s^{-1}) de la nube de iones es directamente proporcional

al campo eléctrico aplicado (E , Vcm^{-1}), siendo K ($\text{cm}^2\text{V}^{-1}\text{s}^{-1}$) la constante de movilidad y Esta se describe a nivel molecular por la ecuación de Mason-Schamp (ecuación Ec 9.2).

$$v_d = K \cdot E \quad \text{Ec 9.1}$$

$$K = \left(\frac{3ez}{16N}\right) \left(\sqrt{\frac{2\pi}{kT}}\right) \left(\sqrt{\frac{m+M}{mM}}\right) \left(\frac{1+\alpha}{\Omega}\right) \quad \text{Ec 9.2}$$

donde e es la carga del electrón (1.602×10^{-19} C), z es la carga del ión, N es la densidad del gas de deriva, k la constante de Boltzmann ($1,381 \times 10^{-23}$ J K^{-1}) y T la temperatura en Kelvin, m y M son la masa del ión y del gas de deriva respectivamente, α es un término de corrección dependiente de la masa y Ω es la sección transversal de colisión, derivada de las integraciones de todos los posibles ángulos de dispersión de la sección transversal de la especie ión-molécula neutra de gas de deriva¹²⁻¹⁴.

La constante de movilidad (K) es función de la densidad de gas y, por tanto, presenta una alta dependencia de la temperatura y de la presión; en consecuencia, usualmente se normaliza a condiciones estándar (273K y 760Torr) según la ecuación Ec 9.3. Esta constante normalizada reportada como movilidad reducida (K_0) es necesaria para comparar datos obtenidos bajo diferentes condiciones experimentales.

$$K_0 = K \left(\frac{273}{T}\right) \left(\frac{P}{760}\right) \quad \text{Ec 9.3}$$

El conocimiento de la relación entre la movilidad K_0 de una especie y su estructura molecular ha sido objeto de estudio en IMS a lo largo de los últimos años para finalmente, poder predecir la movilidad de iones desconocidos sin necesidad de comparar con el valor de K_0 de referencia. Se han publicado estudios para predecir movilidades los cuales utilizan modelos complejos los cuales requieren de descriptores estructurales obtenidos a partir de cálculos teóricos previos^{15,16}.

Recientemente se ha propuesto un método empírico (basado en la linealización de la ecuación de Mason-Schamp), en el cual se representa la inversa de la constante de movilidad en función del número de carbonos para diferentes series homólogas de compuestos polares alifáticos, obteniéndose modelos lineales con una buena capacidad predictiva^{17,18}. Sin embargo, los resultados mostrados se refieren solo a series de moléculas sencillas con un solo grupo funcional por cada serie homóloga.

En este trabajo se propone la aplicación del IMS para la caracterización de los surfactantes no iónicos, del tipo alcohol graso polioxietilenado, como un método simple de aplicar y rápido en la obtención de los resultados.

9.2. MÉTODO EXPERIMENTAL

9.2.1. Muestras

Las muestras usadas para este estudio pertenecen a dos tipos: patrones puros de alcoholes etoxilados con un número definido de grupos etoxi y surfactantes comerciales constituidos por mezclas de oligómeros de un alcohol graso con diferente número de grupos etoxi.

Los patrones utilizados fueron el tetraetilenglicol monododecil éter $\geq 98\%$ GC, hexaetilenglicol monododecil éter $\geq 98\%$ TLC, heptaetilenglicol monododecil éter $\geq 98\%$ TLC, hexaetilenglicol monotetradecil éter $\geq 99\%$ GC, octaetilenglicol monohexadecil éter $\geq 98\%$ GC y hexaetilen estearil éter $\geq 98\%$ GC (Sigma-Aldrich, Alemania).

Los surfactantes industriales utilizados fueron: polietilenglicol monododecil éter con un promedio de grupos etoxi de 4 (Brij L4) y otro de 10; polietilenglicol monohexadecil éter con un promedio de 10 (Brij C10); polietilenglicol monododecil éter de 2 (Brij 52), 10 (Brij S10 y Brij 76) unidades etoxi de promedio (Sigma-Aldrich, Alemania); y el Brij 76 (Fluka Chemie AG, Suiza). Además se ha utilizado polietilenglicol 400 (PEG400) que es una mezcla de oligómeros (Merck, Alemania).

Para el análisis, todas las soluciones se prepararon con isopropanol de grado HPLC $>99.8\%$ (Scharlab, España).

Para la designación de un compuesto único, la nomenclatura utilizada en este estudio es C_nE_m , donde C_n se refiere al número de átomos de carbono alifáticos y E_m se refiere al número de grupo etoxi.

9.2.2. Espectrometría de movilidad iónica

El equipo IONSCAN-LS (Smiths Detection, USA) equipado con una lámina de ^{63}Ni como fuente de ionización radioactiva se usó para separar y detectar los distintos compuestos. Se empleó el software IMStation v5.389 (Smiths Detection, USA) para la adquisición y procesamiento de los datos. El espectro de movilidad iónica se adquirió en modo positivo usando como calibrante interno e ión reactivo la nicotinamida, que tiene una movilidad reducida (K_0) de $1.86\text{cm}^2\text{V}^{-1}\text{s}^{-1}$. El intervalo del ciclo del obturador de iones fue de 0.2ms, valor fijado por el fabricante. El espectro de movilidad se adquirió hasta un tiempo de deriva de 40ms. El campo eléctrico aplicado en el tubo de deriva fue de 252Vcm^{-1} , con un potencial de deriva de 1763V y la longitud del tubo de deriva de 7cm. El flujo de aire seco, fijado en 300mLmin^{-1} , se introdujo como gas de deriva al final de la región de deriva.

Para introducir la muestra se usó la desorción térmica, depositando $1\mu\text{L}$ de la solución en una membrana de Teflón. Esta membrana se calentó para vaporizar el analito, el cual era transferido a

la cámara de ionización. Las temperaturas de desorción, inlet y tubo de deriva fueron 290, 280 y 230°C, respectivamente.

9.3. RESULTADOS Y DISCUSIÓN

9.3.1. Fundamento teórico

La movilidad de una especie viene dada por la ecuación de Michael-Schamp (ecuación Ec 9.2) que, cuando se combina con la ecuación Ec 9.3, se obtiene la ecuación para la movilidad iónica reducida (Ec 9.4).

$$K_0 = \left(\frac{273}{T}\right) \left(\frac{P}{760}\right) \left(\frac{3ez}{16N}\right) \left(\sqrt{\frac{2\pi}{kT}}\right) \left(\sqrt{\frac{m+M}{mM}}\right) \left(\frac{1+\alpha}{\Omega}\right) \quad \text{Ec 9.4}$$

Los 4 primeros términos de la ecuación son constantes bajo las mismas condiciones experimentales de temperatura y presión, y pueden ser sustituidos por la constante $1/\Psi$. Una segunda consideración, es que el término masa reducida [$\mu=(m+M/mM)$] se puede considerar prácticamente constante e igual a M si $m \gg M$ ^{12,18}, para dar la ecuación Ec 9.5 y su inversa, la ecuación Ec 9.6.

$$K_0 = \frac{1}{\psi} \left(\sqrt{\frac{1}{M}}\right) \left(\frac{1+\alpha}{\Omega}\right) \quad \text{Ec 9.5}$$

$$\frac{1}{K_0} = \frac{\psi\sqrt{M}}{1+\alpha} \Omega \quad \text{Ec 9.6}$$

La sección transversal de colisiones Ω es la suma de todos los posibles ángulos de dispersión de la sección transversal de colisiones de una molécula^{19,20} y se puede expresar como la suma de las secciones transversales de colisiones desglosada en diferentes partes (ecuación Ec 9.7), agrupados en fragmentos estructuralmente similares.

$$\frac{1}{K_0} = \sum_{i=1}^n \left(\frac{\psi\sqrt{M}}{1+\alpha}\right) \Omega_i \quad \text{Ec 9.7}$$

En los alcoholes etoxilados, Ω se puede desglosar como suma de Ω_n , debida a la contribución de la cadena alquílica, Ω_m , debida a la contribución de los grupos etoxi de los oligómeros y Ω_c debida a los grupos terminales H- y -OH restantes (Figura 9.1), dando lugar a la ecuación Ec 9.8.

$$\frac{1}{K_0} = \left(\frac{\psi\sqrt{M}}{1+\alpha} \Omega_n \right) + \left(\frac{\psi\sqrt{M}}{1+\alpha} \Omega_m \right) + \left(\frac{\psi\sqrt{M}}{1+\alpha} \Omega_c \right) \quad \text{Ec 9.8}$$

Cada uno de estos términos de Ω_i se puede asociar con cada uno de los tres coeficientes de la Figura 9.1: Ω_n está asociado con ρ (constante de proporcionalidad entre $1/K_0$ y el número de carbonos alifáticos), Ω_m está asociado con κ (constante de proporcionalidad entre $1/K_0$ y el número de grupos etoxi) y Ω_c está asociado con σ (término independiente asociado a los grupos terminales H- y -OH). Por lo tanto, cada uno de los términos de la ecuación Ec 9.8 se puede considerar asociado a una subunidad de un alcohol etoxilado lineal, de forma que cada una de las partes contribuye al total del valor de $1/K_0$ (ecuación Ec 9.9).

$$\frac{1}{K_0} = \sigma + \rho \cdot n + \kappa \cdot m \quad \text{Ec 9.9}$$

Los espectros de movilidad iónica de alcoholes etoxilados muestran una serie de picos a diferentes tiempos de deriva, pertenecientes a los diferentes compuestos (etoxímeros) con diferente número de grupos etoxi; al aumentar el número de grupos etoxi (aumento del peso molecular) aumenta el tiempo de deriva y el valor de $1/K_0$.

Los valores de K_0 se corrigen con respecto a una referencia (nicotinamida) mediante la ecuación Ec 9.10.

$$K_{0(comp)} = K_{0(ref)} \cdot \frac{t_{d(ref)}}{t_{d(comp)}} \quad \text{Ec 9.10}$$

donde $K_{0(comp)}$ es la constante de movilidad reducida del compuesto, $K_{0(ref)}$ es la constante de movilidad del compuesto de referencia, que en este caso es el de la nicotinamida usado como calibrante interno, $t_{d(ref)}$ y $t_{d(comp)}$ son los tiempos de deriva de la nicotinamida y de cada pico del espectro de movilidad iónica, respectivamente.

9.3.2. Caracterización de los espectros de movilidad iónica

El espectro A de la Figura 9.2 muestra el espectro de movilidad iónica de una mezcla de etoxímeros de un surfactante de la serie C₁₂ (polietilenglicol monododecil éter con 10 grupos etoxi de promedio) en el que se numeraron los diferentes picos del 1 al 13. En la misma figura se muestran los espectros de tres patrones puros de monododecil éter con 4, 6 y 7 grupos etoxi. Al comparar estos espectros, se observó que los tiempos de deriva de los picos coincidían respectivamente con los picos 1, 3 y 4 de la mezcla de etoxímeros (espectro A). Esta correlación demostraba que los alcoholes etoxilados aparecen en el espectro ordenados, en tiempos de deriva (y en $1/K_0$) según el número de grupos etoxi, por lo que los compuestos que aparecen en el espectro A numerados del 1 al 13 pudieron ser asignados de 4 hasta 16 grupos etoxi respectivamente.

Este mismo concepto se aplicó a otros surfactantes polioxietilenados con distinta longitud de la cadena hidrocarbonada. Se pudo identificar cada etoxímero de cada una de las mezclas oligoméricas. Los espectros de movilidad de PEG 400 (serie C₀), polietilenglicol monododecil éter (serie C₁₂), polietilenglicol monohexadecil éter (serie C₁₆) y polietilenglicol monooctadecil éter (serie C₁₈) permitieron obtener los valores de las constantes de movilidad reducida (K_0) de todos los picos que aparecían en los mismos. La Tabla 9.1 muestra los valores de $1/K_0$ encontrados para cada compuesto, ordenadas en cuatro columnas según la serie a la cual pertenecen.

En la Figura 9.3 se representan los valores de $1/K_0$ versus el número de grupos etoxi m asignados en la Tabla 9.1. Las cuatro series (C₀, C₁₂, C₁₆ y C₁₈), con diferente longitud de cadena alquílica n , presentaban un buen ajuste lineal para cada una de las rectas. Los diferentes valores de la ordenada en el origen se atribuyeron a la distinta longitud de la cadena hidrocarbonada de los alcoholes grasos. Un gráfico de los valores de ordenada en el origen en función del número de carbonos de la cadena alquílica del surfactante condujo también a una relación lineal. La ecuación obtenida ($Intercept = 0.0346n + 0.3935$, $R^2 = 0.9973$) mostraba el comportamiento de la variación de $1/K_0$ únicamente para la parte alquílica. Los valores de la intercepción correspondían estructuralmente a los valores de $1/K_0$ en ausencia de la parte oligomérica y por lo tanto al equivalente estructural del alcohol primario. La ecuación reportada por Hariharan et al. para la serie de alcoholes primarios ($1/K_0 = 0.0355N_C + 0.4345$, $R^2 = 1.0000$)¹⁶ es muy similar a la ecuación encontrada. Las pendientes de ambas ecuaciones, que indicaban la proporcionalidad entre $1/K_0$ y el número de carbonos de la cadena alquílica, fueron muy semejantes.

Tabla 9.1 Valores promedio de la $1/K_0$ experimental obtenidas de los productos de cada serie homóloga: C_0 (PEG 400), C_{12} (Brij L4 y DeEGMD), C_{16} (Brij C10) y C_{18} (Brij52 y Brij S10); valores calculados a partir de la ecuación obtenida vía MLR y asignación de las unidades etoxi.

C_0 series			C_{12} series			C_{16} series			C_{18} series		
$1/K_0$ Experimental (Vscm⁻²)	$1/K_0$ Calculado (Vscm⁻²)	N° grupos etoxi asignados									
0.6697	0.6729	4	1.0776	1.0692	4	1.2142	1.2008	4	1.2159	1.2011	3
0.7360	0.7393	5	1.1435	1.1347	5	1.2777	1.2663	5	1.2809	1.2666	4
0.8063	0.8057	6	1.2114	1.2002	6	1.3440	1.3318	6	1.3438	1.3321	5
0.8683	0.8721	7	1.2777	1.2657	7	1.4082	1.3973	7	1.4077	1.3976	6
0.9289	0.9385	8	1.3428	1.3312	8	1.4720	1.4628	8	1.4720	1.4631	7
0.9989	1.0049	9	1.4071	1.3967	9	1.5337	1.5283	9	1.5341	1.5286	8
1.0704	1.0713	10	1.4709	1.4622	10	1.5959	1.5938	10	1.5961	1.5941	9
1.1398	1.1377	11	1.5350	1.5277	11	1.6571	1.6593	11	1.6579	1.6596	10
1.2085	1.2041	12	1.5968	1.5932	12	1.7200	1.7248	12	1.7175	1.7251	11
1.2771	1.2705	13	1.6589	1.6587	13	1.7788	1.7903	13	1.7784	1.7906	12
			1.7209	1.7242	14	1.8365	1.8558	14	1.8358	1.8561	13
			1.7765	1.7897	15	1.8972	1.9213	15	1.8939	1.9216	14
						1.9432	1.9868	16			

Del mismo modo, usando patrones del mismo número de grupos etoxilo, $C_{12}E_6$, $C_{14}E_6$ y $C_{18}E_6$, se obtuvo una relación lineal de $1/K_0$ frente al número de átomos de carbono alquílicos; la ecuación obtenida fue $1/K_0=0.0329n+0.8191$. A partir de esta ecuación se podían predecir los valores de $1/K_0$ de todos los compuestos C_nE_6 de cada serie homóloga.

La ecuación Ec 9.11, derivada de la ecuación teórica de Michael-Schamp, permitía calcular la relación entre $1/K_0$ de alcoholes etoxilados y el número de carbonos de la cadena lineal n y el número de grupos etoxi m .

$$\frac{1}{K_0} = \sigma + \rho \cdot n + \kappa \cdot m \quad \text{Ec 9.11}$$

donde los coeficientes ρ y κ son las constantes de proporcionalidad entre $1/K_0$ y el número de carbonos alifáticos y el número de grupos etoxi respectivamente, obtenidas en las ecuaciones anteriores. Por otro lado, σ es la ordenada en el origen de la ecuación que se puede atribuir al resto del compuesto que no pertenece ni a la cadena alquílica ni a la parte oligomérica, es decir, a los grupos terminales (H- y -OH). Estos coeficientes se podían calcular directamente ajustando un modelo de regresión lineal múltiple (MLR), que tiene la ventaja de incluir en el ajuste las posibles influencias de las dos variables (n y m) en la variable predictora ($1/K_0$). Para la construcción del modelo MLR se tomaron como variables el número de carbonos n y el número de grupos etoxi m de la parte oligomérica y como respuesta el valor de $1/K_0$. De los valores de K_0 disponibles, correspondientes a 54 oligómeros, se escogieron 38 de forma aleatoria para formar parte del set de

calibración, procurando que el espacio muestral fuera homogéneo. Los 16 oligómeros restantes se utilizaron como set de predicción.

Los coeficientes obtenidos del modelo fueron 0.0329 (ρ) para el número de carbonos n , 0.0655 (κ) para el número de grupos etoxi m y un término independiente igual a 0.4121 (σ) con un coeficiente de determinación de 0.998 y un bias de $2.65 \cdot 10^{-8}$.

$$1/K_0 = 0.0329 \cdot n + 0.0655 \cdot m + 0.4121 \quad \text{Ec 9.12}$$

Los coeficientes ρ y κ , representan la pendiente de la recta en el espacio tridimensional formado por las dos variables n y m y la variable predictora $1/K_0$. La bondad del modelo se evaluó mediante la representación de los valores experimentales de $1/K_0$ del set de validación frente a los valores predichos por el modelo, obteniendo una recta de pendiente 0.987, offset 0.013 y $R^2=0.999$. Estos valores no son significativamente distintos de 1 y 0 respectivamente, por lo que se pudo concluir que el modelo obtenido tenía una adecuada capacidad predictiva.

El valor del RMSEP fue de 0.006. Un test *t de Student* de los residuales proporcionó un valor de t_{calc} (1.707) menor que la $t_{\text{crítica}}$ (2.042) para un intervalo de confianza del 95%; lo que confirmaba la bondad del modelo lineal propuesto.

Nótese que el valor de κ es aproximadamente el doble que el de ρ ; al aumentar en una unidad m se añaden dos unidades de CH_2 equivalentes a $2n$ de la parte alquílica. Esto indicaba que en un alcohol lineal etoxilado la contribución a la constante de movilidad reducida dependía del número de unidades CH_2 que contenía la molécula. La equivalencia entre dos unidades de CH_2 con un grupo etoxi (con dos CH_2) permitió establecer que el valor de $1/K_0$ estaba directamente relacionado con el número de átomos de carbono y que la contribución de los grupos éter ($-\text{O}-$) era prácticamente nula. El ajuste por mínimos cuadrados de los valores de $1/K_0$ frente al número total de átomos de C del compuesto dio la ecuación $1/K_0=0.0325n+0.4268$; que presentaba un buen valor del ajuste $R^2 = 0,999$. Esta ecuación podía utilizarse para determinar el número total de átomos de C de la cadena a partir de la determinación de $1/K_0$.

Por otro lado, el valor del término independiente encontrado con el modelo MLR, $\sigma=0.4121$, era muy similar al valor de $1/K_0$ del clúster $(\text{H}_2\text{O})\text{H}^+$ reportado por la bibliografía²¹, que es de 0.42. Esto permitió relacionar éste término independiente σ con la contribución que aportan a la $1/K_0$ los grupos terminales de los alcoholes etoxilados lineales ($-\text{H}$ y $-\text{OH}$), que como se ha apuntado anteriormente, es asimismo muy semejante al reportado los alcoholes primarios (0.4345)¹⁶.

9.3.3. Validación del modelo establecido

Se aplicó el modelo obtenido a la caracterización de diferentes surfactantes comerciales industriales, realizando la predicción de la presencia de los diferentes oligómeros en muestras comerciales de surfactantes. Los residuales obtenidos por comparación de los valores de I/K_0 experimentales y calculados con la ecuación obtenida fueron un buen indicativo de la bondad del ajuste.

En los espectros de algunos compuestos aparecían picos no asignables al surfactante, ya que aparecían a tiempos de deriva bajos (indicativos de menor PM) y presentaban un comportamiento distinto a los picos que correspondían al LAE en cuestión. Incluso llegaban a aparecer a tiempos de desorción mayores, todo y su menor peso molecular. Esto indicaba la diferencia de volatilidad entre un grupo y el otro. Estos picos de muy baja amplitud y de menor PM, se atribuyeron a oligómeros de polietilenglicol, formados durante la síntesis de los surfactantes. La intensidad de estos picos aumentaba en los surfactantes que fueron adquiridos unos años antes de este estudio, de lo que se dedució que estos productos se descomponen con el tiempo, dando oligómeros de surfactante de menor índice de etoxilo y polietilenglicol de bajo índice de polimerización.

Con el objeto de establecer la presencia de polietilenglicoles en el compuesto, se analizó una muestra a la que se ha añadido polietilenglicol. La Figura 9.4 muestra los espectros del PEG 400 y de un surfactante comercial (BrijS10) sin adición y con la adición de PEG 400 respectivamente. Puede observarse que los picos del PEG aparecen a tiempos de deriva bajos y que se superponen en parte con los del BrijS10, aumentando su amplitud. Esto indicó que los cinco primeros picos que aparecían en el Brij S10 podían pertenecer a polietilenglicoles (marcados del 1 al 5).

Para contribuir a la diferenciación se desdoblaron los espectros en los diferentes tiempos de desorción tal y como se muestra en la Figura 9.4. No fue posible detectar los oligómeros con un elevado número de grupos etoxilo, lo que se atribuyó a su insuficiente volatilidad; solamente se detectaron oligómeros hasta 16 grupos etoxi, con independencia de que se tratase de surfactantes de C12, C14, C16 y C18. En la Figura 9.5 de la izquierda se muestran los espectros de movilidad del BrijS10 solo y en la de la derecha este mismo alterado con PEG400, para diferentes tiempos de desorción. Se muestra la ausencia (o muy baja concentración) de PEG de bajo PM en el BrijS10, los cuales sí se presentaban en el espectro alterado con PEG. Éste presenta los picos característicos del PEG a bajos tiempos de desorción y de deriva, propios de su bajo PM y mayor volatilidad. Ahora bien, los picos del número 1 al 5 en la Figura 9.5 de la derecha se observa que aparecen en parte a 0.8 y a 1.2s siguiendo la tendencia de desorción en productos que pertenecen a una misma familia; esto es, a mayor peso molecular va disminuyendo la volatilidad. Con esto se demostró que las señales que presenta el Brij S10 entre el 1 y el 5 pertenecían a polietilenglicoles.

La asignación de los picos del Brij S10 (C18En) (Figura 9.4 espectro b) se realizó aplicando la ecuación Ec 9.12. Los residuales (diferencia entre el valor experimental y el predicho por la ecuación) fueron de bajo valor, lo que indicó que la asignación realizada era correcta. En la Tabla 9.2 se muestran los diferentes valores de l/K_0 para los diferentes picos obtenidos del espectro y la asignación realizada de los diferentes oligómeros. Esta asignación se realizó teniendo en cuenta tanto los residuales como el comportamiento de estos analizando el espectro de movilidad. Así, del pico 1 al 5 pertenecían a polietilenglicoles que presentaba el producto Brij S10, y del pico número 6 al 15 a oligómeros de la serie C18.

Tabla 9.2 Señales obtenidas al analizar el Brij S10 (C18Em) alterado con polietilenglicol (C0Em).

Nº pico	Tiempo deriva (ms)	1/K ₀ Experimental (Vscm ⁻²)	Suponiendo n*=0			Suponiendo n*=18		
			m** teórica	1/K ₀ Calculada (Vscm ⁻²)	Residual	m** teórica	1/K ₀ Calculada (Vscm ⁻²)	Residual
1	16.3290	0.9295	8	0.9364	0.0069			
2	17.5490	0.9990	9	1.0019	0.0029			
3	18.7810	1.0691	10	1.0674	-0.0017	1	1.0701	0.0010
4	20.0080	1.1390	11	1.1329	-0.0061	2	1.1356	-0.0034
5	21.1780	1.2055	12	1.1984	-0.0071	3	1.2011	-0.0044
6	22.5070	1.2813	13	1.2639	-0.0174	4	1.2666	-0.0147
7	23.6280	1.3449	14	1.3294	-0.0155	5	1.3321	-0.0128
8	24.7310	1.4078	15	1.3949	-0.0129	6	1.3976	-0.0102
9	25.8830	1.4734	16	1.4604	-0.0130	7	1.4631	-0.0103
10	26.9725	1.5353	17	1.5259	-0.0094	8	1.5286	-0.0067
11	28.0695	1.5978°	18	1.5914	-0.0064	9	1.5941	-0.0037
12	29.1395	1.6586	19	1.6569	-0.0017	10	1.6596	0.0010
13	30.1725	1.7175	20	1.7224	0.0049	11	1.7251	0.0076
14	31.1810	1.7778	21	1.7879	0.0101	12	1.7906	0.0128
15	32.2495	1.8357	22	1.8534	0.0177	13	1.8561	0.0204

* n=número de carbonos de la cadena alquílica; ** m= unidades de grupos etoxi del oligómero.

También se estudió la aplicación de este modelo en surfactantes etoxilados constituidos por una mezcla de alcoholes grasos, con el objeto de establecer la posibilidad de aplicar el modelo a mezclas de surfactantes. Se registraron los espectros de una mezcla de oligómeros de surfactantes de C13 y C15 y se determinaron los valores de $1/K_0$. Paralelamente se calcularon los valores de $1/K_0$ aplicando la ecuación Ec 9.12 para cada C₁₃E_m, C₁₄E_m y un C₁₅E_m individualmente y se calcularon los residuales entre los valores experimentales y los calculados. El gráfico de los residuales frente al número de carbonos de la cadena alquílica (Figura 9.6 derecha) mostró claramente que los oligómeros del C₁₄E_m presentaban residuales más bajos. Esto podía explicarse por la reducida resolución que posee el IMS, que no permitía la diferenciación de los oligómeros de C13 y de C15; y al ser la proporción de alcoholes grasos de partida similar, se obtuvo experimentalmente un espectro aparente que pertenecía al valor medio de carbonos de la cadena alquílica.

El análisis de una mezcla de C₁₆E_m y C₁₈E_m, con una proporción de alcoholes grasos desconocida, condujo a los residuales mostrados en la Figura 9.6 izquierda, donde se ve que el de menor valor corresponde al C₁₆E_m y ligeramente superiores para C₁₈E_m; en este caso los mayores residuales corresponderían al C₁₇E_m.

9.4. CONCLUSIONES

Para los alcoholes lineales etoxilados, la constante de movilidad corregida se relaciona inversamente con la sección transversal de colisiones (Ω), como se deduce de la ecuación de Mason-Schamp, y se puede considerar constituida por tres términos relacionados con las tres subunidades que componen los alcoholes etoxilados: los grupos metileno de la cadena hidrocarbonada, los grupos etoxi de la parte oligomérica y los grupos terminales. Esta hipótesis permitió deducir, usando los valores experimentales de K_0 , una ecuación matemática que incluye el número de átomos de carbono de la cadena alifática y el número de grupos etoxi. La calidad del ajuste obtenido con los datos experimentales demuestra que la hipótesis formulada es correcta. Esta hipótesis permitió establecer la distribución de oligómeros sin necesidad de utilizar patrones.

La ecuación empírica calculada demostró que las dos partes de la molécula se comportan de manera semejante pero independiente, confirmado por el hecho de la gran semejanza del coeficiente de la parte alifática con el descrito en la bibliografía para la serie de alcoholes primarios.

Este estudio demuestra la aplicabilidad del IMS en la caracterización de alcoholes etoxilados lineales de relativamente bajo peso molecular (número de grupos etoxilo <15) así como la detección de los oligómeros de polietilenglicoles normalmente presentes en este tipo de surfactantes. Además de la separación por PM, se produce una separación por volatilidad que facilita la detección y caracterización de los productos industriales.

9.5. BIBLIOGRAFÍA

1. Strege, M. A. Total residue analysis of swabs by ion mobility spectrometry. *Anal. Chem.* **81**, 4576–

- 4580 (2009).
2. Qin, C., Granger, A., Papov, V., McCaffrey, J. & Norwood, D. L. Quantitative determination of residual active pharmaceutical ingredients and intermediates on equipment surfaces by ion mobility spectrometry. *J. Pharm. Biomed. Anal.* (2010).
 3. Armenta, S., Alcalà, M., Blanco, M. & González, J. M. Ion mobility spectrometry for the simultaneous determination of diacetyl midecamycin and detergents in cleaning validation. *J. Pharm. Biomed. Anal.* (2013).
 4. Strege, M. A., Kozerski, J., Juarbe, N. & Mahoney, P. At-Line Quantitative Ion Mobility Spectrometry for Direct Analysis of Swabs for Pharmaceutical Manufacturing Equipment Cleaning Verification. *Anal. Chem.* **80**, 3040–3044 (2008).
 5. Cassani, G., Pratesi, C., Faccetti, L. & Pravettoni, S. Characterization of Alcohol Ethoxylates as Alcohol Ethoxy Sulfate Derivatives by Liquid Chromatography – Mass Spectrometry. *J. Surfactants Deterg.* **7**, 195–202 (2004).
 6. Komarek, K. & Kostrubanicova, E. Capillary gas chromatography – mass spectrometry of lower oxyethylenated aliphatic alcohols 1. *J. Chromatogr. A* **773**, 219–226 (1997).
 7. Micó-Tormos, A., Simó-Alfonso, E. F. & Ramis-Ramos, G. Single-pump bi-dimensional LC applied to the characterization of derivatized fatty alcohol ethoxylates. *J. Sep. Sci.* **33**, 1398–404 (2010).
 8. Rudewicz, P. & Munson, B. Analysis of Complex Mixtures of Ethoxylated Alcohols by Probe Distillation/Chemical Ionization Mass Spectrometry. *Anal. Chem.* **58**, 674–679 (1986).
 9. Asmussen, C. & Stan, H.-J. Determination of Non-Ionic Surfactants of the Alcohol Polyethoxylate Type by Means of High Temperature Gas Chromatography and Atomic Emission Detection. *J. High Resolut. Chromatogr.* **21**, 597–604 (1998).
 10. Zu, C., Praay, H. N., Bell, B. M. & Redwine, O. D. Derivatization of fatty alcohol ethoxylate non-ionic surfactants using 2-sulfobenzoic anhydride for characterization by liquid chromatography / mass spectrometry. *Rapid Commun. mass Spectrom.* **24**, 120–128 (2010).
 11. Micó-Tormos, a, Bianchi, F., Simó-Alfonso, E. F. & Ramis-Ramos, G. Determination of fatty alcohol ethoxylates with diphenic anhydride derivatization and liquid chromatography with spectrophotometric detection: a comparative study with other anhydrides. *J. Chromatogr. A* **1216**, 3023–30 (2009).
 12. Borsdorf, H. & Eiceman, G. A. Ion Mobility Spectrometry: Principles and Applications. *Appl. Spectrosc. Rev.* **41**, 323–375 (2006).
 13. G.A.Eiceman; Z.Karpas. *Ion mobility spectrometry*. (CRC Press Taylor and Francis, 2005).
 14. Creaser, C. S. *et al.* Ion mobility spectrometry : a review . Part 1 . Structural analysis by mobility measurement. *Analyst* **129**, 984–994 (2004).
 15. Wessel, M. D. & Jurs, P. C. Structural Information Using Multiple Linear Regression Analysis and Computational Neural Networks. *Anal. Chem.* **86**, 2480–2487 (1994).
 16. Wessel, M. D., Sutter, J. M. & Jurs, P. C. Prediction of Reduced Ion Mobility Constants of Organic Compounds from Molecular Structure monitor (EVM) gas chromatography / ion mobility spec-. *Anal. Chem.* **68**, 4237–4243 (1996).
 17. Hariharan, C., Ingo Baumbach, J. & Vautz, W. Empirical prediction of reduced ion mobilities of secondary alcohols. *Int. J. Ion Mobil. Spectrom.* **12**, 59–63 (2009).
 18. Hariharan, C. B., Baumbach, I. & Vautz, W. Linearized Equations for the Reduced Ion Mobilities of Polar Aliphatic Organic Compounds. *Anal. Chem.* **82**, 427–431 (2010).

19. Louis, R. H., Hill, H. H. & Eiceman, G. A. Ion Mobility Spectrometry in Analytical Chemistry. *Crit. Rev. Anal. Chem.* **21**, 321–355 (1990).
20. Masson, E. A. & Mcdaniel, E. W. *TRANSPORT PROPERTIES*. (1988).
21. Kim, S. H., Betty, K. R. & Karasek, F. W. Mobility Behavior and Composition of Hydrated Positive Reactant Ions in Plasma Chromatography with Nitrogen Carrier Gas. *Anal. Chem.* **50**, 2006–2012 (1978).

10. CUANTIFICACIÓN DE ALCOHOLES GRASOS LINEALES ETOXILADOS POR ESPECTROMETRÍA DE MOVILIDAD IÓNICA (IMS)

10.1. INTRODUCCIÓN

En el capítulo previo se ha estudiado la aplicación de IMS en la caracterización de los alcoholes lineales etoxilados (LAE) encontrando la ecuación empírica Ec 10.1, que relaciona la inversa de la constante de movilidad reducida (K_0) con la longitud de la cadena hidrocarbonada, con el número de grupos etoxi de la parte oligomérica y los grupos terminales –H y –OH.

$$\frac{1}{K_0} = 0.033 \cdot n + 0.065 \cdot m + 0.412 \quad \text{Ec 10.1}$$

En el presente trabajo se estudia un método para determinar la distribución de los diferentes oligómeros de una serie homóloga y se propone un método para la determinación cuantitativa de cada una de ellos individualmente que componen el tensioactivo.

10.2. MÉTODO EXPERIMENTAL

10.2.1. Muestras

Las muestras usadas para este estudio pertenecen a dos tipos. (1) Patrones puros de alcoholes etoxilados con un número definido de grupos etoxi. Es necesario señalar que estos patrones no son 100% puros (contienen pequeñas cantidades de impurezas, otros oligómeros y humedad), lo que los hace muy útiles en la identificación de oligómeros pero no totalmente adecuados para el análisis cuantitativo. (2) Surfactantes comerciales constituidas por mezclas de oligómeros de un alcohol graso con diferente número de grupos etoxi.

Los patrones utilizados fueron el tetraetilenglicol monododecil éter $\geq 98\%$ GC, hexaetilenglicol monododecil éter $\geq 98\%$ TLC, heptaetilen glicol monododecil éter $\geq 98\%$ TLC, hexaetilen glicol monotetradecil éter $\geq 99\%$ GC y hexaetilen estearil éter $\geq 98\%$ GC (Sigma-Aldrich, Alemania).

Los surfactantes industriales usados fueron el polietilenglicol monododecil éter con promedios de grupos etoxi de 4 (Brij L4), de 10 (DeEGMD) y de 23 (Brij L23); polietilenglicol monohexadecil éter con un promedio de 10 (Brij C10) y de 20 (Brij C20); polietilenglicol monooctadecil éter con promedios de 2 (Brij S2 y Brij 52) y de 10 (Brij S10) (Sigma-Aldrich, Alemania); y el Brij 76 (Fluka Chemie AG, Suiza). Además se utilizó polietilenglicol 400 (PEG400) que es una mezcla de oligómeros (Merck, Alemania).

Todas las soluciones se prepararon con isopropanol grado HPLC >99.8% (Scharlab, España). Para los patrones se prepararon disoluciones en un rango de concentración entre 0.2 y 10 mg/L y para los surfactantes comerciales se han preparado disoluciones entre 5 y 200 mg/L.

Para la designación de un compuesto único, la nomenclatura utilizada en este estudio es C_nE_m , donde C_n se refiere al número de átomos de carbono alifáticos y E_m se refiere al número de grupos etoxi.

10.2.2. Espectrometría de movilidad iónica

El equipo IONSCAN-LS (Smiths Detection, USA) equipado con una lámina de ^{63}Ni como fuente de ionización radioactiva se usó para separar y detectar los distintos compuestos. Se empleó el software IMStation v5.389 (Smiths Detection, USA) para la adquisición y procesamiento de los datos. El espectro de movilidad iónica se adquirió en modo positivo usando como calibrante interno e ión reactivo la nicotinamida, que tiene una movilidad reducida (K_0) de $1.86\text{cm}^2\text{V}^{-1}\text{s}^{-1}$. El intervalo del ciclo del obturador de iones fue de 0.2ms, valor fijado por el fabricante. El espectro de movilidad se adquirió hasta un tiempo de deriva de 40ms. El campo eléctrico aplicado en el tubo de deriva fue de 252Vcm^{-1} , con un potencial de deriva de 1763V y la longitud del tubo de deriva de 7cm. El flujo de aire seco, fijado en 300mLmin^{-1} , se introdujo como gas de deriva al final de la región de deriva.

Para introducir la muestra se usó la desorción térmica, depositando un micro litro de la solución en una membrana de Teflón. Esta membrana se calentó para vaporizar el analito, el cual era transferido a la cámara de ionización. Las temperaturas de desorción, inlet y tubo de deriva fueron 300, 300 y 230°C respectivamente.

10.2.3. Procesado de datos

El software MATLAB v7.11.0 (The MathWorks, USA) fue el empleado para procesar los datos.

El equipo IMS proporciona un cubo de datos donde la primera dimensión es el tiempo de deriva, la segunda el tiempo de análisis o de desorción y por última, la intensidad (amplitud). Este cubo inicial se importó a Matlab para procesar los datos. Se eliminaron los primeros 13 milisegundos del tiempo de deriva ya que solo aparece el pico correspondiente al reactivo ionizante. Por otro lado, también se seleccionaron solo los 4 primeros segundos de desorción ya que el resto solo aporta ruido e información no relevante.

Se ensayaron dos filtros distintos para reducir el ruido: el filtro de Fourier que consiste en transformar los datos del dominio tiempo al dominio frecuencia mediante el algoritmo de transformada rápida de Fourier (FFT por sus siglas en inglés *Fast Fourier Transform*), que elimina

las frecuencias altas, las que pertenecen al ruido, eligiendo una frecuencia de corte determinada y finalmente se aplica la inversa de FFT para volver a obtener los datos en dominio tiempo. El otro filtro ensayado fue el filtro de Savitzky-Golay mediante la aplicación de un polinomio de grado 3 y una ventana móvil de 21 puntos.

En ambos casos, con los datos filtrados, se calculó el promedio obteniendo por un lado un gráfico de las intensidades frente al tiempo de deriva convencional, y por otro lado, se construyó un gráfico de superficie de tiempo de deriva frente tiempo de desorción donde los contornos representaban las intensidades (en unidades arbitrarias) de los diferentes componentes.

10.2.4. GC-MS

El análisis se llevó a cabo con un cromatógrafo de gases Agilent 6890 GC equipado con un detector de masas cuadrupolar Agilent MS 5973 (Agilent Technologies, USA). La columna usada fue una HP-5MS (30m x 0.25mm), que es una columna capilar de sílice fundida recubierta con 0.25 μ m de poli(fenilmetil)siloxano como fase estacionaria. Como gas portador se usó helio a un flujo constante de 1ml/min. La técnica de inyección empleada fue la *splitless* y el inyector se mantuvo a 325°C. La temperatura de la columna inicialmente se programó a 100°C, manteniéndose durante 2 minutos. Después se aumentó de 100 a 325°C a una velocidad de 4°C/min. Finalmente, se mantuvo a 325°C durante 35min. Las temperaturas de la línea de transferencia y de la fuente de iones fueron 280°C y 230°C respectivamente. El detector de masas funcionó con el modo de ionización electrónica con un potencial de ionización de 70eV.

10.3. RESULTADOS Y DISCUSIÓN

10.3.1. Establecimiento de parámetros de análisis y tratamiento de datos

Usando disoluciones de patrones puros se estableció la respuesta del IMS con la concentración para los diferentes tipos de surfactantes; los resultados muestran una respuesta lineal, entre 0,2 y 1 mg/L, con la concentración. En la Tabla 10.1 se muestran los parámetros de las rectas de calibración. Las diferencias que se observan entre las pendientes y las ordenadas en el origen se atribuyen principalmente al distinto grado de pureza de los patrones (y diferente contenido en humedad) y en menor grado al error experimental de la medida. Se puede considerar que todos los compuestos tienen una respuesta similar (independiente del peso molecular o de la longitud de la cadena de grupos etoxi) en el detector del espectrómetro de movilidad iónica, lo que aplicaremos en el estudio para determinar la distribución de los diferentes oligómeros.

Tabla 10.1 Figuras de mérito de las rectas de calibración de diferentes surfactantes etoxilados en el rango de concentraciones entre 0.2 y 1mg/L,

Patrón	Pendiente	Ordenada en el origen	R ²
C ₁₂ E ₄	19.166	17.293	0.9947
C ₁₂ E ₆	22.819	12.189	0.9973
C ₁₂ E ₇	21.814	11.705	0.9976
C ₁₄ E ₆	21.360	13.675	0.9687

Algunas variables experimentales del espectrómetro IM están directamente relacionadas a la respuesta obtenida al analizar alcoholes lineales etoxilados y se evaluaron individualmente para optimizar sus valores. La temperatura de desorción puede ser un factor importante en el análisis de este tipo de compuestos, ya que determina por un lado la volatilidad y por otro la estabilidad térmica de estas sustancias; se realizó un estudio de la respuesta del surfactante comercial DeEGMD (C₁₂E_m) en un rango de temperatura desorción entre 200 y 300°C que es la temperatura máxima de trabajo del instrumento utilizado. En la Figura 10.1 se presentan los espectros de movilidad iónica obtenidos a temperatura 230, 280 y 300°C. Al aumentar la temperatura, incrementa el número de oligómeros detectados de mayor peso molecular, sin que se observe la aparición de picos de oligómeros de menor peso molecular. Esto indica que la desorción aumenta con la temperatura sin que se produzca degradación térmica de los mismos. El resto del estudio se realizó a la temperatura de 300°C, que es la máxima que permite el instrumento.

También se llevó a cabo un estudio de la concentración máxima de surfactante que se podía determinar, considerando que el factor limitante en esta modalidad de IMS es la concentración de reactivo ionizante. La cantidad de muestra introducida no debe consumir todo el reactivo ionizante para evitar la reducción de respuesta por falta de ionización, lo que es particularmente frecuente en el análisis de las mezclas. El rango de concentraciones de trabajo de los surfactantes se estableció

entre 10 y 100 mg/L, puesto que a concentraciones más altas se consumía completamente el reactivo ionizante.

Para el análisis de los surfactantes comerciales se prepararon disoluciones entre 10 y 100 mg/L y se inyectaron 1 μ L de cada disolución por triplicado; obteniéndose un cubo de datos para cada producto. El instrumento IMS proporciona un cubo de datos de intensidad de señal para cada fila (tiempo de deriva) y cada columna (tiempo de análisis o desorción). En la Figura 10.2a se muestra gráficamente el cubo de datos para el DeEGMD. Además, se hizo el *defolding* por filas y por columnas del cubo, tal como se muestra en la Figura 10.2b y Figura 10.2c. Con ellas, se pudo observar que en los tiempos de deriva comprendidos entre 0 y 11ms aparece la señal correspondiente al reactivo ionizante. Por otro lado, se vio que los tiempos de análisis altos, a partir de los 4 segundos, no presentaban señal (solo aportaban ruido), por lo que en los siguientes tratamientos se eliminaron esas partes de los datos. También se puso de manifiesto que la señal presentaba un elevado ruido por lo que se decidió aplicar un filtro adecuado para mejorarla.

Se ensayaron filtros Fourier con diferentes frecuencias de corte, pero con ninguna de ellas se encontró una relación correcta entre la reducción del ruido y la atenuación de la señal. Otro filtro ensayado fue el filtro de Savitzky-Golay aplicando distintos grados de polinomio y valores de ventana móvil; los mejores resultados fueron con un polinomio de tercer grado y usando una ventana de 21 puntos. En la Figura 10.3 se muestra la comparación del promedio de la matriz

original y el mismo promedio una vez aplicado el filtro de Fourier y de Savitzky-Golay; se puede observar la mayor eficacia del segundo filtro, que no produce prácticamente atenuación sustancial de la señal mientras que hay una reducción significativa del ruido.

Con el objeto de mejorar la señal del espectro de movilidad iónica y como paso previo al estudio cuantitativo, se propuso reducir el cubo de datos, eliminando los tiempos largos de análisis y los tiempos de deriva cortos. Seguido de aplicar un filtro de Savitzky-Golay de polinomio de grado 3 y ventana móvil de 21 puntos para reducir el ruido.

La Figura 10.4 muestra el gráfico de contorno (donde el eje de abscisas es el tiempo de análisis y el de ordenadas el tiempo de deriva) del cubo de datos reducido de: a) el patrón $C_{12}E_7$, b) el producto Brij L4 ($C_{12}E_m$) y c) el polietilenglicol con un promedio de peso molecular de 400 (C_0E_m). La Figura 10.4b muestra dos grupos de picos en el eje del tiempo de análisis. Se puede observar que para un mismo tiempo de deriva aparecen dos señales a distintos tiempos de desorción, esto es, productos con peso molecular similar tienen puntos de ebullición distintos. Al comparar este gráfico con el correspondiente al patrón $C_{12}E_7$, se determinó que los picos de tiempo de análisis menor correspondían a los oligómeros $C_{12}E_m$. Nótese que dentro del mismo grupo, al aumentar los tiempos de deriva, y por lo tanto el peso molecular, los tiempos de desorción aumentan ligeramente pero siguen perteneciendo al $C_{12}E_m$. Por el contrario, el segundo grupo de picos, al compararse con la Figura 10.4c, se determinó que pertenecían a los polietilenglicoles que frecuentemente acompañan a los surfactantes. Por lo tanto, era necesario considerar que los picos de PEG de mayor PM (mayor tiempo de deriva) podían solaparse con los picos que presentara el surfactante $C_{12}E_m$ de menor PM. Esto afectaría a las medidas de la intensidad y por tanto darían resultados erróneos en el cálculo del promedio de grupos etoxilo de los oligómeros. Para cada compuesto se podía establecer

un intervalo de tiempos de análisis para hacer el promedio de etoxilos, de manera que incluyera todas las señales pertenecientes a los LAE y a la vez que excluyera a las pertenecientes a PEG.

10.3.2. Distribución de los oligómeros

Asumiendo que la respuesta del detector era la misma para los diferentes oligómeros (como se ha demostrado anteriormente), se determinó la señal (amplitud) de cada oligómero en cada surfactante. Se ha calculado el porcentaje que representaba la amplitud de cada oligómero respecto a la suma total de amplitudes sin tener en cuenta los picos de PEG. En el Figura 10.5 se muestran algunos de los histogramas construidos a partir de las amplitudes relativas calculadas. Estas se correspondían a la distribución de los diferentes oligómeros LAE que formaban parte del tensioactivo. Se observaron claramente dos familias distintas. Por un lado la distribución que mostraban los productos Brij S2 (Figura 10.5b), Brij 52 y Brij L4, los cuales mostraban un primer pico mayoritario y que disminuía la intensidad al aumentar las unidades de grupos etoxi. Por otro lado, la que presentaban el resto de productos, DeEGMD (Figura 10.5c), Brij L23, Brij C10, Brij C20 y Brij 76; con una distribución con un máximo semejante a la observada en la bibliografía reportada en la introducción de este artículo.

A partir de la distribución de los diferentes oligómeros se pudo calcular el número medio de grupos etoxi (n) de cada mezcla empleando los porcentajes obtenidos de cada surfactante. En la Tabla 10.2 se recogen estos valores calculados y los indicados por las casas comerciales para cada uno de los productos. Para $n < 10$, los valores experimentales fueron semejantes a los reportados por las especificaciones; pero cuando $n > 10$ los valores calculados fueron mucho menores que el indicado por la casa comercial. Este comportamiento se atribuyó a que los oligómeros de mayor peso molecular no se desorbían a la temperatura de trabajo debido a su menor volatilidad¹. Además, tampoco se observó la presencia de oligómeros de menor peso molecular que pudieran proceder de la pérdida de óxidos de etileno, ni de PEG de bajo peso molecular. Ambas posibles causas condujeron a un cálculo incorrecto de la de la distribución de los oligómeros de los alcoholes etoxilados con una mayor abundancia de los de menor peso molecular. Esto indicó que un IMS con desorción térmico como sistema de introducción de la muestra solo era adecuado para analizar alcoholes etoxilados lineales de mayor volatilidad o, lo que es lo mismo, de peso molecular relativamente bajo.

Tabla 10.2 Valores promedio de grupos etoxilo de diferentes surfactantes comerciales. Los nominales están en las especificaciones comerciales y en las columnas de IMS y GC-MS se indican los valores calculados experimentalmente con ambas técnicas.

Compuesto	Nominal	IMS	GC-MS
	n	n	n
Brij L4	4	6.5	-
DeEGMD	10	8.6	6.8
Brij L23	23	8.5	6.9
Brij 52	2	4.8	-
Brij S10	10	9.8	-
Brij 76	10	8.5	-
Brij C10	10	8.3	-
Brij C20	20	6.3	-

Con el objetivo de validar los resultados obtenidos con el IMS, se analizó un producto representativo de cada familia mediante la caracterización sin derivatización por GC-MS. Se utilizó la cromatografía de gases porque, como ocurre en el IMS, el compuesto debe ser volátil a la temperatura de trabajo para poder ser analizados. Por otro lado, la espectrometría de masas permite detectar todos los oligómeros presentes. Este método está ampliamente descrito en la bibliografía.

En concreto se analizaron los patrones y las mezclas DeEGMD ($C_{12}E_n$), BrijL23 ($C_{12}E_n$) y Brij S2 ($C_{18}E_n$). Para simplificar los cromatogramas se aplicó el detector selectivo de iones (SIM por sus siglas en inglés *Single Ion Monitoring*), seleccionando el pico de masa molecular que corresponde al fragmento del alcohol con un único grupo etoxi. Como pasa en IMS, en GC, para una misma

serie homóloga, el tiempo de retención aumenta linealmente con el peso molecular. Los patrones se utilizaron para identificar los LAE. Para establecer la distribución de los diferentes oligómeros se supuso que la respuesta de cada oligómero en SIM era la misma. A partir de las áreas de cada pico, se determinó la distribución de oligómeros presentes en cada producto comercial analizado por GC-MS, obteniendo como resultado un histograma. En la Figura 10.6 se muestran los correspondientes al producto DeEGMD y al Brij S2. Presentaban ciertas diferencias con las distribuciones obtenidas en IMS. En el caso del DeEGMD (Figura 10.6a), aunque seguían una misma distribución, en el caso de gases, el máximo lo presentaba a un oligómero menor. Además, el producto de mayor peso molecular observado fue el $C_{12}E_{11}$, en cambio en IMS era el $C_{12}E_{16}$. Estas observaciones experimentales se podrían explicar por el hecho que en cromatografía de gases, tal como está recogido en la bibliografía, los LAE de alto peso molecular son inestables térmicamente. Esto es, los oligómeros con cadena etoxilada más larga se romperían y aumentaría aparentemente la presencia (y en consecuencia el área) de los oligómeros de menor peso molecular. Este mismo comportamiento se observó en el BrijS2. Para cada tensioactivo también se calculó la media de óxidos de etileno presentes. Los resultados se encuentran en la cuarta columna de la Tabla 10.2. Como en el caso del IMS cuando m era igual o mayor a 10, los promedios obtenidos en cromatografía de gases fueron menores que los teóricos. Esto se explica, como se encontró en la bibliografía, por la baja volatilidad de los LAEs de mayor peso molecular y por la inestabilidad térmica de los mismos. Todo esto verificó los resultados obtenidos con el análisis en IMS.

10.3.3. Estudio cuantitativo de los oligómeros

Como paso previo a la cuantificación, se llevó a cabo un estudio para determinar el efecto que tenían en la señal de un LAE la presencia de otros oligómeros. Para ello se analizaron distintas soluciones de una misma concentración (4mg/L) del patrón $C_{12}E_4$, mientras que en cada una de ellas se añadieron distintas cantidades (entre 0 y 10 mg/L) de otro estándar, en este caso el $C_{12}E_7$, con el objeto de ver si la amplitud de la señal del $C_{12}E_4$ variaba al aumentar la concentración de

$C_{12}E_7$. En la Figura 10.7 se muestra como la amplitud de la señal para una misma concentración de $C_{12}E_4$ va disminuyendo a medida que aumenta la presencia de $C_{12}E_7$. Cada punto representa el promedio de tres inyecciones y la barra de error representa la desviación estándar obtenida de estas. Con ello se determinó que la señal de cada oligómero era función de la presencia del otro, dada la competición por captar el protón del reactivo ionizante. Este hecho implicaba que no se podía realizar una calibración única para todos los alcoholes lineales etoxilado, ya que dependería de la presencia del resto de oligómeros en cada caso. Era necesario hacer una cuantificación propia para cada producto en concreto.

Debido al efecto que tienen los LAEs entre ellos, se propuso la cuantificación de estos mediante una modificación de la adición estándar generalizada. Se hizo una simplificación porque era muy complicado disponer de patrones de todos los oligómeros presentes en una mezcla. Se estudió el DeEGMD como modelo. La determinación individual de cada uno de los oligómeros que formaban parte de este tensioactivo se realizó añadiendo cantidades conocidas de los patrones $C_{12}E_4$, $C_{12}E_6$ y $C_{12}E_7$ a una disolución de concentración total conocida de DeEGMD, tal como se ve en la Tabla 10.3. En la Figura 10.8 se presenta el gráfico de histogramas de las intensidades obtenidas para cada LAE en las disoluciones analizadas. Las muestras modificadas presentaban un aumento en la intensidad del oligómero añadido a la mezcla. Al saber la concentración exacta que se añadió de patrón en cada muestra, este incremento se pudo relacionar directamente con la cantidad de patrón presente en la muestra. Esto permitió obtener una constante de proporcionalidad para cada caso. Con el promedio de esta relación y con las intensidades de los diferentes picos que presentaba una muestra de 9.80mg/L de DeEGMD, se calculó la concentración individual de cada LAE. La suma de las concentraciones calculadas individualmente debería dar el valor de la concentración total de muestra preparada. El valor obtenido experimentalmente fue 9.67mg/L. Esto representó, en términos de exactitud, un error relativo del 1.3%. La repetitividad de este método se evaluó analizando tres veces seguidas la misma muestra y calculando la desviación estándar relativa

(RSD) de la concentración total encontrada con la misma relación señal/ruido. Se obtuvo un valor de 6% el cual podría ser comparable al reportado por Asmussen et al. que era de un rango entre 2 y 4 % analizando sustancias de referencia².

Tabla 10.3 Diseño de disoluciones de DeEGMD adicionadas con distintas cantidades de patrón.

Muestra	C ₁₂ E _m (mg/L)	C ₁₂ E ₄ (mg/L)	C ₁₂ E ₆ (mg/L)	C ₁₂ E ₇ (mg/L)
1	10	0	0	0
2	10	0.6	0	0
3	10	0	0.6	0
4	10	0	0	0.8
5	10	0.4	0.8	0.8

10.4. CONCLUSIONES

Las amplitudes de las señales de los espectros de movilidad permitieron obtener distribuciones satisfactorias para mezclas con un promedio de grupos etoxi menor a 10. Se encontró que los oligómeros de peso molecular alto no eran desorbidos. El método era comparable al GC-MS. Además, se encontró que en IMS los LAEs compiten entre sí para captar el protón del reactivo ionizante. La cuantificación se realizó de forma individual para cada producto comercial. Esta se hace mediante la simplificación de una adición estándar generalizada y usando una constante de proporcionalidad para calcular la concentración de cada oligómero presente en la muestra. Los valores obtenidos fueron satisfactorios y la repetitividad del método era comparable con la bibliografía.

Con todo esto, este estudio aportó un método rápido y de bajo coste para caracterizar y cuantificar mezclas de alcoholes lineales etoxilados con promedio de grupos etoxi igual o menor a diez mediante la técnica IMS.

10.5. BIBLIOGRAFÍA

1. Micó-Tormos, A., Simó-Alfonso, E. F. & Ramis-Ramos, G. Single-pump bi-dimensional LC applied to the characterization of derivatized fatty alcohol ethoxylates. *J. Sep. Sci.* **33**, 1398–404 (2010).
2. Asmussen, C. & Stan, H.-J. Determination of Non-Ionic Surfactants of the Alcohol Polyethoxylate Type by Means of High Temperature Gas Chromatography and Atomic Emission Detection. *J. High Resolut. Chromatogr.* **21**, 597–604 (1998).

Parte V

ANÁLISIS DATOS DE PROCESO

11. REVISIÓN HISTÓRICA DE LOS DATOS DE PROCESO

11.1. INTRODUCCIÓN

La fabricación de un lote de un medicamento autorizado genera una gran cantidad de datos que quedan almacenados en diferentes bases de datos, dependiendo del origen de los mismos. Se entiende como datos aquellos que están relacionados con la calidad de los materiales de partida que se han utilizado para su producción, los parámetros de proceso de cada una de las operaciones unitarias por las que ha pasado con el fin de obtener el producto final, todos los parámetros de calidad de los productos intermedios y finales, incidencias o desviaciones ocurridas durante todo el proceso, etc. Todos estos datos son una fuente muy valiosa de información que no siempre es explotada debido a la dificultad de organizar y manejar tanta cantidad de datos.

Los parámetros de calidad de los materiales de partida y productos finales se deben evaluar anualmente como parte del informe *Product Quality Review* (PQR), que sirve para verificar la correcta calidad del medicamento a lo largo de todo su ciclo de vida –fabricación, estabilidad y entono¹. Como parte del PQR, se debe demostrar que el proceso está bajo control estadístico, es decir, que la variabilidad del proceso es debida a causas comunes o intrínsecas y no a causas asignables. Este análisis se suele llevar a cabo mediante el análisis univariable de los parámetros críticos de calidad del producto (CQA). Los CQA se evalúan individualmente con el control estadístico del proceso (SPC por sus siglas en inglés *Statistical Process Control*). La herramienta empleada son los gráficos de control, donde se establecen unos límites de control a partir de datos históricos. Los datos que pertenecen al periodo anual en estudio se representan en estos gráficos. Si algún punto sigue algunas de las reglas de Nelson más sensibles, se considera que con una probabilidad significativa, el proceso está fuera de control y es necesario encontrar la causa raíz que provoca este hecho². Ahora bien, las variables pueden estar correlacionadas y esto queda ocultado realizando un análisis univariable. Es decir, puede que individualmente el proceso está aparentemente bajo control, pero que si se analizaran las CQA de forma conjunta, se pondría de manifiesto que realmente el proceso no está bajo control, o también se puede dar el caso contrario, un proceso aparentemente fuera de control estadístico que si se analizan todas las variables conjuntas el proceso es estable³. Es bajo estas premisas que apareció el concepto de control estadístico multivariable (MSPC por sus siglas en inglés *Multiple Statistical Process Control*)⁴. El objetivo es establecer un modelo PCA con el cual se puedan detectar muestras no comunes durante la producción⁵. Los límites de control definidos son los valores de Hotelling T^2 y los residuales Q obtenidos durante la calibración del modelo usando lotes históricos “normales”, la variabilidad es debida solo a causas intrínsecas del proceso.

El estudio de datos históricos también es el primer paso cuando se pretende variar la fabricación tradicional de un medicamento autorizado a una fabricación bajo el concepto de QbD⁶⁻⁸. Estos se pueden usar para encontrar las relaciones empíricas de los CMA, CPP y CQA que permitan definir rangos de los factores, y si no fueran suficientes, se puede proponer DoE para establecer un espacio de diseño. Además, los datos históricos también se pueden usar para determinar cuál es el estado de control de los procesos de producción e identificar carencias o potenciales puntos de mejora del proceso. Para este fin, se puede realizar un estudio de las trayectorias de los CPP de los lotes, mediante un MPCA (*multiway PCA*)⁹ dado que las CPP son variantes con el tiempo. También podría servir para monitorizar el proceso durante la producción de nuevos lotes y detectar en tiempo real si ocurren fallos¹⁰⁻¹². Una de las mayores dificultades para realizar un análisis de las trayectorias de la fabricación de los lotes es que la duración no es siempre igual, el tiempo puede variar lote a lote. Así, una de las primeras cosas a realizar es un escalado del tiempo que permita alinear todos los lotes. Una manera de alinear el tiempo de todos los lotes es transformarlo a una escala común usando la interpolación lineal¹³.

El objetivo de este trabajo fue analizar los datos históricos de un medicamento presentado en forma de comprimidos recubiertos que sirviera para evaluar la implementación del MSPC como herramienta para demostrar el estado de control del proceso respecto a las CQA y a las CPP.

11.2. MÉTODO EXPERIMENTAL

11.2.1. Obtención de los datos y preparación de las matrices

La formulación del comprimido recubierto consta de un 10.3% de API y 5 excipientes, la función de cada uno de ellos está detallada en la Tabla 11.1. La fabricación de un lote de este producto farmacéutico implica 4 operaciones unitarias. En este trabajo solo se tuvieron en cuenta las que tenían un impacto en la transformación del producto. Estas operaciones son las que se llevan a cabo en el equipo granulador de lecho fluido, mezcla de los componentes iniciales de la formulación, granulación y secado; la compresión, que se realiza en dos máquinas indistintamente; y finalmente el recubrimiento. El acondicionamiento se descartó del estudio debido a que no tiene ningún impacto en los parámetros de calidad evaluados en este trabajo. Se recopilaron los datos de calidad de materiales de partida, los parámetros de proceso de cada operación unitaria y los datos de calidad de los productos intermedios y finales. En la Tabla 11.1 se definen cada uno de ellos. El periodo de tiempo que abarcaban los datos era de 3 años. A continuación se explican las matrices de cada estudio.

Tabla 11.1 Parámetros de calidad de los materiales de partida, parámetros del proceso y parámetros de calidad del producto final.

Parámetros de proceso		Parámetros de calidad del producto acabado	
Operación unitaria	Parámetro	Producto	Parámetro
Lecho fluido	Temperatura del deshumidificador	Comprimido	Dureza media (kp)
	Temperatura de entrada		Masa (mg/tab)
	Temperatura del producto		Coefficiente variación de la masa (%)
	Temperatura de salida	Comprimido recubierto	Dureza media (kp)
	Caudal de entrada		Grosor (mm)
	Presión de pulverización		Masa (mg/tab)
	Velocidad de la bomba peristáltica		Coefficiente variación de la masa (%)
	Porcentaje de extracción		Contenido agua (%)
Compresión	Presión previa	Isómero R (-) (%)	
	Presión principal	Riqueza API (%)	
	Velocidad del alimentador	Ensayo disolución (%)	
	Velocidad del rotor	Uniformidad (desviación estándar)	
Recubrimiento	Temperatura de entrada		Uniformidad (valor de aceptación)
	Temperatura de salida		
	Caudal de entrada		
	Humedad relativa de salida		
	Presión de pulverización		
	Velocidad de la bomba peristáltica		
	Velocidad del contenedor		
Vacío interior			

11.2.1.1. MSPC

El estudio MSPC se realizó con una matriz que contenía la información de los 13 CQA del producto intermedio -el comprimido sin lacar- y del producto acabado de 511 lotes fabricados en un periodo de tres años. La dimensión de la matriz era de 511×13 (*lotes x variables*). Todos los lotes incluidos fueron analizados por parte del departamento de control de calidad. De los 511 lotes, hubo 13 que tenían algún parámetro de calidad que estaba fuera de los límites de especificación internos. Todos los demás lotes habían sido aprobados para la liberación al mercado de los productos.

11.2.1.2. Modelado de la trayectoria de proceso

Se recopilaron las variables de proceso de 27 lotes históricos. Estas variables correspondían a las 3 operaciones unitarias bajo estudio. Así, para el lecho fluido, se tenía una matriz de $(27 \text{muestras} \times \text{tiempo}) \times 8 \text{ variables}$. Para cada uno de los equipos de compresión, la matriz era de $(27 \text{muestras} \times \text{tiempo}) \times 4 \text{ variables}$. Por último, las dimensiones de la matriz correspondiente al recubrimiento eran $(27 \text{muestras} \times \text{tiempo}) \times 8 \text{ variables}$.

Dado que el tiempo que duraba cada lote en cada una de las operaciones unitarias no era el mismo, los datos iniciales se tuvieron que transformar para alinear las trayectorias. Cada matriz inicial, como ya se ha comentado, tenía una dimensión de $(\text{lote} \times \text{tiempo}) \times \text{variable de proceso}$, siendo el tiempo variable para cada lote. El método de alineamiento que se usó fue la interpolación lineal¹³. Para este fin, se usó un programa existente desarrollado con Python (Python Software Foundation), que se encuentra en el repositorio (<https://github.com/krlosMata/csv-parser>). Lo que realizaba dicho programa era separar la matriz inicial por productos, normalizar el tiempo a la unidad e interpolar linealmente las variables de proceso. Una vez obtenidos los archivos individuales para cada lote, con el mismo número de filas (tiempo), estos se cargaron en el Matlab v7.11.0 (The MathWorks, USA) con el fin de construir la matriz 3D necesaria para realizar los análisis posteriores (estructura en la Figura 11.1).

11.2.2. Análisis de los datos

11.2.2.1. MSPC de parámetros de calidad

El análisis multivariable de los datos se llevó a cabo con el software Solo 8.1 (Eigenvector Research, USA). Además, se realizó un análisis para detectar valores atípicos con el test de Grubbs para el cual se empleó el software Minitab 17 (Minitab, USA). Se utilizaron los lotes de los dos primeros años para desarrollar el modelo PCA, un total de 357 muestras. De estas, se escogieron 195 (el 60% del total) mediante el algoritmo de Kennard-Stone para realizar la calibración del modelo. Se aseguró que los lotes incluidos no hubieran presentado incidencias, que los parámetros de calidad se encontraran dentro de los límites de especificación y que representaran toda la variabilidad esperable. Una vez aplicado el auto-escalado (centrado y escalado) de los datos por variables, se calculó el modelo PCA. La validación cruzada se realizó con la ventana veneciana usando 100 subconjuntos de 3 muestras cada uno de ellos. El número de componentes principales se escogió de acuerdo a la varianza explicada, el RMSECV y procurando que todas las variables (CQA) quedaran explicadas¹⁴. Con este modelo se fijaron los límites de control con los valores de Hotelling T^2 y los residuales Q con un 99% de nivel de confianza. El resto de los lotes de desarrollo de modelo, 162, se emplearon para optimizar y evaluar el modelo. Se evaluó la capacidad del modelo de identificar correctamente los lotes que tiene una variabilidad esperable, así como detectar aquellos lotes que tienen una variabilidad diferente al resto¹⁵.

El modelo MSPC calibrado y validado se aplicó a los lotes del tercer año, 155 muestras. La matriz de datos se proyectó en el modelo PCA. Se evaluaron los gráficos de *scores* y de los límites T^2 y residuales Q para determinar la idoneidad del modelo y su estado de control estadístico, respectivamente.

En todos los casos anteriores, si se detectaba algún lote fuera de los límites de Hotelling T^2 y/o residuales Q, se empleó el gráfico de contribuciones para determinar qué variable o variables eran responsables de provocar la inusual variación.

11.2.2.1. MSPC de parámetros de proceso

El análisis multivariable de los datos se llevó a cabo con el software Solo 8.1 (Eigenvector Research, USA). Dado que los parámetros de proceso varían con el tiempo, no era posible aplicar un análisis con PCA, sino que se usó un algoritmo que permitía reducir la información de manera multivariable, con el MPCA. Se usó la matriz de datos estructurada como se muestra en la Figura 11.1 y el software la descompuso como se muestra en la Figura 11.2.

El modelo MPCA se calculó usando 13 muestras de producción que presentaban una trayectoria considerada común. El número de componentes principales que formaron parte del modelo se escogieron en base a la varianza explicada total. Con este modelo se fijaron los límites de control con los valores de Hotelling T^2 y los residuales Q con un 99% de nivel de confianza. Para la validación del modelo se proyectaron 14 muestras. De estas, cuatro presentaban la misma trayectoria que las de calibración, las cuales se usarían para determinar la capacidad del modelo para clasificar correctamente muestras que eran parecidas al target; mientras que 10 tenían una trayectoria diferente y se usarían para evaluar la capacidad del modelo y de los límites fijados para detectar lotes con una evolución del proceso distinta.

11.3. RESULTADOS Y DISCUSIÓN

11.3.1. MSPC de los parámetros de calidad

El análisis estadístico descriptivo inicial a las 357 muestras sirvió para detectar lotes que presentaban alguna variabilidad fuera de lo común. Entre ellos se detectaron 4 lotes con resultados de algún parámetro fuera de especificaciones. Uno presentaba una dureza mayor, dos una riqueza por encima y por debajo de los límites y uno no cumplía el test de uniformidad de contenido. Además, para cada una de las variables que se detectaron valores sospechosos –relacionados con la masa del núcleo y el producto final, sus coeficientes de variación relacionados, el contenido en agua y el ensayo de disolución- por lo que se realizó un test de Grubbs. Se confirmó que había un total de 7 lotes con valores atípicos, dado que el resultado del test rechazaba la hipótesis nula (valor de $p < 0.05$). Estas muestras se incluyeron directamente en el conjunto de validación del modelo. El resto de lotes, los cuales se consideraron que tenían un rango normal, se dividieron en el conjunto de calibración (60% del total) y validación (40%) usando el algoritmo Kennard Stone. Las características de los CQA de los lotes que se usaron para la calibración del modelo PCA se presentan en la Tabla 11.2. En ella se describe el rango de los valores incluidos en el conjunto de calibración en cada una de las variables de calidad. Los datos con mayor variabilidad son la dureza, de los núcleos y los comprimidos recubiertos, el contenido en agua y el valor de aceptación de la uniformidad de contenido.

Tabla 11.2 Características de los datos empleados para desarrollar el modelo PCA para definir los límites de control del MSPC.

	Dureza producto intermedio (kp)	Masa producto intermedio (mg/tab)	CV masa producto intermedio (%)	Dureza (kp)	Grosor (mm)	Masa (mg/tab)	CV masa (%)	Contenido agua (%)	Isómero R (-) (%)	Riqueza API (%)	Ensayo disolución (%)	Uniformidad (desv. Est.)	Uniformidad (valr. Accept.)
Valor máximo	7.50	262.40	2.50	7.30	3.50	266.40	2.90	5.00	0.49	102.78	104.00	5.10	12.30
Valor mínimo	5.60	259.20	1.00	4.50	3.30	260.40	0.90	0.33	0.16	95.00	91.40	0.21	0.50
Promedio	6.71	260.20	NA	5.66	3.47	263.84	NA	3.71	0.33	98.63	99.09	NA	3.21
Desv. Estándar	0.4	0.3	NA	0.5	0.0	0.9	NA	0.6	0.1	1.6	2.1	NA	2.7
CV (%)	6.3	0.1	NA	8.6	1.4	0.3	NA	17.0	21.7	1.6	2.1	NA	85.6

El modelo PCA se calculó incluyendo 3 componentes. Estos se escogieron por los siguientes motivos. El RMSECV, aunque no varía de forma significativa de un componente a otro, con el tercer componente es con el que se obtiene un valor menor con una varianza explicada del 53.7%, la cual se considera un valor suficiente en este tipo de datos (Tabla 11.3). Pero la razón más importante fue que con los tres primeros componentes, se incluía información de todos los CQA, tal como se demuestra con el gráfico de la varianza capturada por cada PC en la Figura 11.3. En el caso del contenido en agua la variación es capturada prácticamente solo en el tercer PC, por lo que un número menor de componentes principales para calcular el modelo PCA no queda justificado. En cambio, un número mayor de componentes incluiría información redundante al modelo.

Tabla 11.3 Varianza explicada en cada componente principal del modelo PCA y los valores de RMSEC y RMSECV calculados.

Número PC	Varianza explicada (%)	Varianza explicada acumulada (%)	RMSEC	RMSECV
1	24.5	24.5	0.87	0.95
2	15.1	39.6	0.78	0.95
3	14.1	53.7	0.68	0.90
4	8.8	62.5	0.61	0.94
5	7.9	70.4	0.54	0.96
6	7.2	77.6	0.47	0.94

Los gráficos de los *loadings* de la Figura 11.4 permitieron establecer correlaciones entre variables. Como era de esperar, los valores relacionados con la uniformidad de contenido presentaban cierta correlación con el coeficiente de variación relacionado con la masa del núcleo y del producto recubierto, de acuerdo al primer PC. En el segundo componente, se observó que el parámetro Q del ensayo de disolución está correlacionado con la dureza del producto y la riqueza del mismo, lo cual tiene lógica. Ahora bien, en el caso del tercer componente, donde la mayor varianza capturada se relaciona al contenido de agua y la presencia de isómero R (-), los *loadings* muestran una

correlación inversa entre ambos CQA. Esta relación, a priori, parece que es casualidad y no causalidad.

El gráfico de *scores* del primer componente principal puso de manifiesto que existía cierta tendencia temporal a partir de la segunda mitad del segundo año (Figura 11.5). A pesar de ello se establecieron los límites de control ya que se consideró que esta variabilidad se había incluido en el conjunto de calibración. Al saberse de antemano que todos los lotes usados para el cálculo del modelo se encontraban dentro del rango normal de trabajo, los límites de Hotelling T^2 y del residual Q se establecieron en un nivel de confianza del 99%, para evitar tener un exceso de alarmas injustificadas.

La validación del modelo se realizó proyectando el conjunto de validación en el modelo PCA calculado. Las muestras que están marcadas en color rosa en la Figura 11.6 presentaron unos valores de T^2 , Q o la combinación de ambos elevados. Un valor elevado del residual Q indica que hay información de la muestra no explicada por el modelo. Un valor elevado de T^2 indica que la muestra tiene una diferencia sistemática respecto a la variabilidad modelada. Los lotes 203, 204, 206 y 207 coincidían con los descartados inicialmente por presentar datos atípicos según el test de Grubbs. Los gráficos de contribución a cada uno de los valores que sobrepasaba los límites, confirmaron que la variable que tenía un valor atípico era la que contribuía más a obtener dicho valor. Ver a modo de ejemplo el gráfico de las muestras 206 y 207 (en la Figura 11.7), donde el valor atípico era el contenido en agua (0.33%) y la masa del núcleo (262.4mg/comprimido), respectivamente. En el resto de muestras, no había una sola variable que contribuyera a la presencia de estos valores fuera de los límites de control, sino que era una combinación de varias variables. Los límites de control establecidos durante la calibración del modelo resultaron ser adecuados para detectar lotes fuera del rango normal y evitar a la vez sobre-reaccionar ante la presencia de alarmas injustificadas.

El análisis de los lotes correspondientes al tercer año sirvió para poner de manifiesto la tendencia temporal detectada en el desarrollo del modelo que persistía con las muestras del nuevo año, demostrando que había habido algún cambio (de materiales de partida, de proceso, analítico...) que no se tenía controlado. La tendencia detectada (Figura 11.5) y el elevado número de muestras fuera de control, un 19%, indicaban que el proceso estaba fuera de control estadístico. La tendencia en el primer componente principal da la información que esta es debida a la uniformidad de contenido/coeficiente de variación de la masa o a la dureza, según la varianza capturada por este factor. Al revisar de manera individual cada una de las variables mencionadas, se detecta que el valor de los parámetros relacionados con la uniformidad de contenido y el coeficiente de variación de la masa han aumentado con el tiempo. Por lo que una vez detectada la causa raíz que lo ha provocado y si es necesario, aplicadas las acciones correctoras, el modelo para aplicar MSPC deberá rehacerse o bien introduciendo más variabilidad o bien eliminando las muestras de antes del cambio. Es importante hacer notar que un proceso fuera de control no significa que el producto esté fuera de las especificaciones, por lo que es apto para el mercado.

11.3.2. Análisis de datos de proceso

La granulación de un producto en un lecho fluido consta principalmente de dos etapas, la de granulación con la pulverización de la solución ligante en el interior del equipo y la del secado del producto para eliminar el exceso de solvente. Asegurarse que los valores de los parámetros de proceso asociados a ambas etapas se encuentran en el intervalo de consigna es clave para tener un producto de la calidad esperada. Además, es importante conocer el estado de control estadístico del proceso de fabricación.

La duración de fabricación de un mismo producto no era siempre la misma entre los lotes. Esta diferencia era debida a distintas razones. La primera fase del proceso consistía en alcanzar una temperatura del producto concreta, por lo que el tiempo que se tardaba en llegar al punto

mencionado no era siempre igual. También se daban las circunstancias de que el proceso se paraba en medio de alguna fase por motivos diversos. Por último, si los datos informativos referentes a un lote quedaban almacenados en el PLC del lecho fluido una vez finalizado el proceso, el aplicativo donde se almacenan los datos de proceso continuaba registrando los parámetros de proceso y quedaban asociados al lote mencionado, sin que hubiera información real relacionada con su producción. Todas las circunstancias mencionadas provocaron que la alineación de los datos de proceso fuera un paso imprescindible antes de su análisis. En este caso, se normalizó el tiempo y se usó la interpolación lineal de las variables para igualar la longitud de los vectores. Es importante remarcar que la transformación realizada no fue adecuada para solventar las dificultades presentadas cuando el registro de parámetros de proceso había continuado una vez finalizada la granulación y secado del producto. Ver la comparación de un lote con una duración de tiempo estándar y la de un lote con un registro de datos extenso en la Figura 11.8. Los lotes que se encontraban en la segunda situación no se emplearon para el desarrollo del modelo, sino como validación externa del mismo para evaluar la capacidad del modelo para detectar lotes con trayectorias distintas.

Una vez alineados los datos, se seleccionaron aquellos lotes que presentaban una trayectoria similar para usarlos como lotes estándares para calcular y validar el modelo MPCA. Además, con el fin de evaluar la capacidad del modelo para detectar procesos con trayectorias totalmente distintas al proceso común se emplearon los lotes de larga duración, mencionados anteriormente, y se simuló lotes con el proceso truncado. Esto es, se eliminó la parte final de los mismos. La calibración se realizó una vez aplicado el auto-escalado por bloques de variables (cada bloque corresponde a un lote) como pretratamiento, y se seleccionaron 2 PC, los cuales explicaban el 59% de la variabilidad de los datos. Este porcentaje de varianza explicada es la habitual en el análisis de parámetros de calidad y de proceso. En la Figura 11.10 se compara la trayectoria de la variable “Temperatura de salida” con los *loadings* del primer componente principal para el mismo

parámetro. El perfil muestra cierta similitud, por lo que parece que prácticamente toda la variabilidad de dicho parámetro se explica con el primer componente. En otros casos, es el segundo componente el que explica la variabilidad o bien la combinación de ambos. En conclusión, los *loadings* servían para relacionar cada variable con el tiempo, por lo que podrían ser útiles para el control durante el proceso. Pero en esta ocasión, se estaba estudiando la capacidad del modelo en diferenciar lotes con comportamientos dentro de los parámetros normales de aquellos que tienen trayectorias distintas. Para ello, se empleó el gráfico de residuales Q frente a los valores T^2 . Se fijaron los límites de control con un 99% de nivel de confianza y se proyectaron las muestras de validación. En la Figura 11.9 se muestran los resultados. Los cuatro lotes considerados correctos (número 17, 22, 29 y 30) cayeron dentro de los límites, mientras que el resto se encontraban fuera de los límites. El modelo era capaz de detectar aquellos procesos que han tenido incidencias o presentan anomalías respecto a aquellos considerados estándares.

Este proceso se repitió para el recubrimiento, obteniendo los mismos resultados satisfactorios y las mismas limitaciones, que era la alineación del tiempo en los casos de lotes con una presencia de datos prolongada una vez finalizado el recubrimiento de las muestras.

En el caso de la compresión, no se consideró necesario aplicar un estudio MPCA a los datos. La razón fue que los parámetros de proceso no dependen del tiempo. Como se puede observar en la Figura 11.11 los parámetros están alrededor de una consigna si el proceso está en marcha. Si se detiene en algún momento, las variables tienen un valor de 0. Se consideró que lo más eficiente era aplicar un gráfico de control basado en PCA. Siempre que el proceso estuviera marcha se obtendría un valor de Q y T^2 que se compararían con los límites calculados durante la calibración, y permitiría establecer si el proceso estaba bajo control estadístico o no.

11.4. CONCLUSIONES

Se ha desarrollado un método MSPC para evaluar el estado de control estadístico de la calidad final de un producto comprimido recubierto. Los límites de control estadístico son los valores de Hotelling T^2 y residual Q fijados con el cálculo de un modelo PCA usando lotes de producción históricos que estaban dentro del rango normal. Este método ha sido capaz de detectar muestras que presentaban una variabilidad distinta a las de calibración, o lo que es lo mismo, tenían una variabilidad asignable, y se ha podido determinar qué factor o conjunto de ellos contribuyen a que una muestra esté fuera de control estadístico. Además se ha demostrado la utilidad del mismo para detectar tendencias en los resultados.

De la misma manera, se ha desarrollado un método MSPC para evaluar las trayectorias de los procesos de producción con el empleo del algoritmo MPCA, que permite el análisis en matrices de 3 dimensiones. Ha sido necesario realizar una alineación de los datos, dado que no siempre tienen la misma duración. Los límites de control han demostrado ser capaces de clasificar como correctos

todos aquellos lotes que presentaban trayectorias similares y ha descartado aquellos que presentaban una distinta. Este método puede emplearse en el PQR para demostrar el estado de control estadístico de proceso, aunque de momento tiene una limitación. La alineación de datos empleada en este trabajo no es apropiada para aquellos lotes que tienen una cantidad de datos altos una vez finalizada la fabricación, por lo tanto es necesario optimizar esta transformación previa. También queda pendiente calcular los límites de control de los *loadings* para poder detectar en tiempo real fallos durante la producción. Este tipo de análisis solo tienen sentido en aquellos procesos que varían con el tiempo, como es el proceso de granulación. Para aquellos que no tienen una dependencia con el tiempo, como por ejemplo la compresión, donde los parámetros están fijados a una misma consigna durante todo el proceso, el empleo de los gráficos de control basados en PCA parece ser más eficiente. Finalmente, el siguiente paso será realizar un modelo PLS de los *scores* obtenidos en los dos modelos MPCA, el del modelo PCA de la compresión y los parámetros de calidad de los materiales de partida con los parámetros de calidad del producto final para encontrar relaciones empíricas entre ellos.

11.5. BIBLIOGRAFÍA

1. AEMPS. Guía de Normas de Correcta Fabricación de Medicamentos de Uso Humano y Veterinario: Anexo 15 Cualificación y validación. (2015).
2. Peng, D. Y. (FDA). Using control charts to evaluate process variability. in *PQRI 2015 Annual Meeting* (2015).
3. Martin, E. B., Morris, A. J., Papazoglou, M. C. & Kiparissides, C. Batch process monitoring for consistent production. *Comput. Chem. Eng.* **20**, S599–S604 (1996).
4. MacGregor, J. F. & Kourti, T. Statistical process control of multivariate processes. *Control Eng. Pract.* **3**, 403–414 (1995).
5. Kjeldahl, K. & Bro, R. Some common misunderstandings in chemometrics. *J. Chemom.* **24**, 558–564 (2010).
6. Schindwein, W. S. & Gibson, M. *Pharmaceutical Quality by Design. A practical Approach.* (2018).
7. Yu, L. X. *et al.* Understanding Pharmaceutical Quality by Design. *AAPS J.* **16**, 771–783 (2014).
8. Yu, L. X. Pharmaceutical quality by design: product and process development, understanding, and control. *Pharm. Res.* **25**, 781–91 (2008).
9. Chen, J. & Liu, J. Derivation of function space analysis based PCA control charts for batch process monitoring. *Chem. Eng. Sci.* **56**, 3289–3304 (2001).
10. Macgregor, J. & Nomikos, P. Multivariate SPC Charts for Batch Monitoring Processes. *Technometrics* **37**, 41–59 (1995).
11. Neogi, D. & Schlags, C. E. Multivariate statistical analysis of an emulsion batch process. *Ind. Eng. Chem. Res.* **37**, 3971–3979 (1998).
12. Yoo, C. K., Lee, J. M., Vanrolleghem, P. A. & Lee, I. B. On-line monitoring of batch processes using multiway independent component analysis. *Chemom. Intell. Lab. Syst.* **71**, 151–163 (2004).
13. Neogi, D. & Schlags, C. E. Multivariate statistical analysis of an emulsion batch process. *Ind. Eng.*

Chem. Res. **37**, 3971–3979 (1998).

14. Tôrres, A. R., de Oliveira, A. D. P., Grangeiro, S. & Fragoso, W. D. Multivariate statistical process control in annual pharmaceutical product review. *J. Process Control* **69**, 97–102 (2018).
15. Liesum, L., Kumpli, D. S., Peinado, A. & McDowall, N. Chapter 16: The Role of Multivariate Statistical Process Control in the Pharma Industry. in *Multivariate Analysis in the Pharmaceutical Industry* 357–384 (Elsevier Inc., 2018). doi:10.1016/b978-0-12-811065-2.00015-1

CONCLUSIONES

CONCLUSIONES

El desarrollo de esta Tesis Doctoral en el ámbito de la industria farmacéutica ha servido para presentar soluciones que permitan progresar hacia una fabricación avanzada de productos farmacéuticos con el uso de dos conjuntos de herramientas PAT. La primera son instrumentos analíticos que dan una respuesta del producto o del estado del proceso, como son el NIR, el Raman y el IMS, y la segunda es el análisis multivariable de los datos. Para ello, se han realizado los siguientes trabajos:

- Se han actualizado y validado bibliotecas espectrales NIR para la identificación y cualificación de 224 materiales de partida y productos granel. Se ha desarrollado un método para usar el Raman como alternativa a la identificación de materia prima.
- Se ha implementado industrialmente un sistema PAT que permite la liberación en tiempo real de un producto granulado mediante el análisis por NIR hacia la etapa de acondicionamiento. Para ello se ha desarrollado y validado un método cualitativo (basado en el coeficiente de correlación) que permite la identificación del producto y dos modelos cuantitativos (basados en PLS) para determinar la concentración de API y el contenido en agua, respectivamente. Estos modelos han demostrado ser robustos para su aplicación en rutina. Por otro lado, también se ha contribuido a la industrialización del sistema participando en las pruebas de validación del sistema y en toda la documentación que conlleva la implementación en rutina de un nuevo método.
- Se ha realizado un primer estudio de viabilidad de un instrumento NIR con sonda multipunto para poder ser implementado en la etapa de dosificación de producto granulados, polvo y comprimidos. Se han determinado las ventajas y limitaciones del mismo. Este estudio forma parte de un nuevo proyecto PAT.
- Se han demostrado las ventajas del IMS para su uso en la validación y verificación de la limpieza como método específico para detectar la presencia de detergentes. Además, se ha demostrado la utilidad de realizar un diseño de experimentos para obtener el conocimiento necesario en aquellos casos que se utilice un método no específico para detecta agentes d elimpieza, como es la conductividad.
- El análisis multivariable de los datos históricos de productos existentes ha puesto de manifiesto la gran información de la que dispone la compañía para obtener conocimiento de los procesos de producción y de los productos. Esta información se podría usar como base para realizar un cambio de paradigma en la fabricación, pasar de fabricar de la forma tradicional a una fabricación bajo QbD.

CONCLUSIONES

Un objetivo muy importante del programa de Doctorats Industrials es la transferencia de conocimiento de la Universidad a la Industria. Durante el transcurso de esta Tesis Doctoral se han llevado a cabo distintas formaciones, tanto a nivel interno como corporativo, sobre el uso de la quimiometria para desarrollar métodos NIR y/o Raman demostrando el cumplimiento de este último propósito.