Control of Raw Materials in the Pharmaceutical Industry by Raman Spectroscopy: Qualification and Development of a Database

Raquel Agostinho
Instituto Superior Técnico, Lisboa, Portugal
May 2018

Abstract

The recent increase in the work volume at Sofarimex has triggered the need to reduce the tests carried out in the quality control laboratory, namely in the verification of raw materials (RM). Thus, a portable equipment using Raman spectroscopy was acquired, that transports the laboratory to the RM sampling room. In order to operate with equipment, a qualification process has been developed and implemented. For this reason, the work focuses on the various stages of equipment qualification, moving to the practical case of portable Raman qualification, from the establishment of the URS (User Requirements Specifications), to the Risk Assessment, where the tests are defined to be performed in Design (DQ), Installation (IQ), Operational (OQ) and Performance (PQ) Qualification. After qualification, it was necessary to develop a database used as reference to establish the Hit Quality Index between the RM spectra to analyze and the reference RM. Consisting of around 400 RM from different suppliers and packaging, the database was validated and successfully tested both in the laboratory and in the warehouse.

Keywords: Equipment Qualification, Raman Spectroscopy, Sampling, Reference Database, Hit Quality Index, Validation.

1. Introduction

The pharmaceutical industry inserted on Europe requires high standards of quality management in the development, manufacture and control of medicines. [1] The regulatory body responsible for regulating pharmaceutical activity in Portugal is the National Authority for Medicines and Health Products I.P. (INFARMED, IP). This entity is responsible for assessing the compliance with regulatory legislation and for the emission of the compliance certificate with GMP (Good Manufacturing Practices) standards. [2] [3] Quality Assurance is responsible for ensuring that products or services meet customer quality requirements and meet regulatory requirements for products or services. [4]

The principles of qualification and validation applicable to the equipment, services and utilities are described in Annex 15 of Volume 4 of Eudralex. [5] In order to ensure that the equipment is properly qualified, it is necessary to consider the critical aspects of the equipment throughout its life cycle. If any changes occur in the equipment that may have an impact on product quality, it should be formally documented, and the impact should be controlled or mitigated. The concept of Quality Risk Management was released in the ICH Q9 guideline[6].
This guideline was launched with the aim of providing guidance to the pharmaceutical industry on a systematic approach to risk management, considering GMPs and quality requirements. This point is important for the equipment qualification process.

**Raman Spectroscopy**

The introduction of PAT (Process Analytical Tool) plays an important role for analyzing risks throughout the life cycle of a pharmaceutical product. Raman spectroscopy is a PAT tool that has been demonstrating scientific value for the control of both process and RM at an early stage of the process. [7]

The phenomenon of inelastic dispersion of light was postulated in 1923 by Smekal, but it was only in 1928 that it was first observed experimentally by Raman and Krishnan. After that moment, the phenomenon has become known as “Raman Effect”. [8] When light interacts with matter, the photons that make up the light can be absorbed, scattered or emitted will no interacts with the material and pass directly through it. If the energy of an incident photon corresponds to the energy gap between the ground state of a molecule and the excited state, the photon can be absorbed, and the molecule is promoted to a higher energy state.[9]

Radiation is often characterized by its wavelength (λ). However, in spectroscopy we are interested in the interaction of radiation with the molecule, so it is analyzed and discussed in terms of energy. Thus, it is common to analyze energy based on frequency scales or wave number. These scales are linearly related to energy, and can be represented mathematically by:

\[ E = h \cdot v \]  

Where \( h \) is Planck constant (\( h = 6.626 \times 10^{-34} \) J.s) and \( v \) is the frequency of light. The velocity of light in vacuum \( c \) (\( c = 2.9979 \times 10^8 \) m.s\(^{-1}\)) is related to the wavelength (\( \lambda \)) by the following formula:

\[ v = \frac{c}{\lambda} \]  

The energy of the electromagnetic waves is directly proportional to the reciprocal wavelength. In the case of vibrational spectroscopy, the reciprocal wavelength is represented by \( k \) and corresponds to the wave number:

\[ E = h \cdot c \cdot k \]  

Usually the wave number \( k \) is given by:

\[ k = \frac{v}{c} = \frac{1}{\lambda} \]  

The equations above show the proportionality relationship between energy and wavelength. Thus, the region of greatest energy is to the left, in figure 1 and the largest wavelength to the right. [11]

In Raman scattering, light interacts with the molecule and distorts (polarizes) the electronic cloud around the nuclei to form an intermediate state, called the “virtual state.” The virtual state is characterized by being neither stable nor long-lasting, so that the photon is rapidly radiated back. The energy changes detected in the vibrational spectroscopy are those
necessary to cause nuclear movement. If only the electron cloud distortion is involved in the dispersion, the photons are scattered with very small frequency changes, since the electrons are comparatively light. [9]

Figure 2 – Stokes Scattered, Rayleigh Scattered e Anti-Stokes Scattered. [9]

The energy difference between the incident and the dispersed photons is represented by the arrows of different lengths. Numerically, the energy difference between the initial and final vibrational states, \( \nu \), or the Raman shift in wave numbers (cm\(^{-1}\)), is calculated by equation 5:

\[
\nu = \left( \frac{1}{\lambda_{\text{incident}}} - \frac{1}{\lambda_{\text{scattered}}} \right) \times 10^7
\]  

(5)

Where \( \lambda_{\text{incident}} \) and \( \lambda_{\text{scattered}} \) are the wavelengths (nm) of the incident photons and the scattered photons, respectively. This is the Raman scatter.

**Instrumentation**

Raman spectroscopy instrumentation has become more commercial and robust with the development of stable laser sources and low noise Charge Coupled Device (CCD) detectors. [12]

These equipments make use of software of control easily parameterized and generally integrated with reference libraries that facilitate the process of identification of raw materials. The most modern instrumentation comprises three basic components: Laser; Optical sampling; Detector. Since inelastic dispersion is a weak phenomenon, Raman instrumentation uses a laser as the source of lighting, because it is a monochromatic source of high-intensity light. Although the laser wavelength can range from UV to near-IR (\( \lambda = 200-1064 \text{nm} \)), most pharmaceutical or bioprocessing applications use near-IR wavelengths (\( \lambda = 785 \text{ or } 830 \text{ or } 1064 \text{ nm} \)), mainly to minimize fluorescence interference, which is the main limitation of Raman spectroscopy.

With a dynamic regulatory scenario and new scientific challenges, Raman spectroscopy has taken proportions as PAT technology in pharmaceutical production due to its wide applicability. [7]

**Raman Spectroscopy in Pharmaceutical Analysis**

Raman Spectroscopy is beneficial for many types of quantitative and qualitative analyzes in several fields, including fundamental chemical research, process control, forensic science and pharmaceuticals. In this way, it can be used as a non-invasive, non-destructive and even non-contact monitoring technology. [13]

Raman spectroscopy can then be used to monitor chemical reactions. The disappearance of the starting materials and the appearance of the products cause easily foreseeable changes in the Raman spectrum. These changes can be used to determine how close the reaction is to the end. Nothing needs to be removed or added to the test tube and data collection and presentation of the results can be completed in just a few seconds. [14]

Raman spectroscopy at-line can control critical process parameters, make process corrections
in real time and ensure the consistent production of the product concerned. New applications for the synthesis of API (Active Principal Ingredient) and crystallization, continuous production and new developments in Raman spectroscopy for understanding and controlling bioprocesses. [15]

**Equipment Qualification**

The Qualification Master Plan is an internal management procedure of Sofarimex, whose main objective is to establish methods and guidelines to implement a qualification program for equipment, facilities and services. The purpose of the qualification process is to ensure the correct suitability, installation, operation and performance of the equipment. The way to prove this is to carry out the activities and tests described in the following documents:

- Risk Assessment (RA);
- Qualification Master Plan;
- Design, Installation, Operational and Performance Protocols;
- Design, Operational and Performance Reports;
- Final Acceptance Report.

**Risk Assessment**

The risk assessment follows the FMECA Strategy (Failure Mode, Effects, and Criticality Analysis). The classification is made according to the following table:

<table>
<thead>
<tr>
<th>Severity</th>
<th>Probability</th>
<th>Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>S=3</td>
<td>P=3</td>
<td>D=3</td>
</tr>
<tr>
<td>Medium</td>
<td>Medium</td>
<td>Medium</td>
</tr>
<tr>
<td>S=2</td>
<td>P=2</td>
<td>D=2</td>
</tr>
<tr>
<td>Low</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>S=1</td>
<td>P=1</td>
<td>D=1</td>
</tr>
</tbody>
</table>

**Table 1 - Calculation of risk level by classifying the RA parameters according to 3 levels.**

![Figure 3 - Steps and Parameters of Risk Assessment used in this study.](image)

2. **Materials and Methods**

**Equipment**

The equipment used in this work is the Bravo Handheld, model from Bruker. This portable Raman Spectrometer is designed for the verification and identification of materials. It is intended for indoor use only, particularly in laboratories and warehouses. There are two different types of measuring tips, which are interchangeable. These measuring tips can be used to measure liquid and solid samples contained in a plastic bottle or bag. With this equipment the samples are measured without direct contact. The equipment has an automated wave number calibration to ensure highly reproducible measurements and a wide spectral range between 3200cm\(^{-1}\) and 300cm\(^{-1}\).

Laser excitation is achieved with a dual laser (700-1100 nm) and due to the fluorescence mitigation functionality, it is possible to identify in seconds dark, fluorescent and weak Raman
signal samples. The equipment complies with the regulations European Pharmacopeia 2.2.48 and USP 1120. [16] The data were stored and compared with the reference spectra using the OPUS software, an indispensable tool for this thesis.

Reference Standards
The standards are: paracetamol in bag, paracetamol in glass vial, benzonitrile in vial, naphthalene in vial, cyclohexane in vial and polystyrene in vial.

3. Results of Raman Qualification
A document with the User Requirements and Specifications (URS) was prepared for the acquisition of the equipment. After that, a Master Plan Qualification was elaborated to define the qualification strategy. Based on the two previous documents a Risk Analysis (26 parameters) was elaborated to establish the tests to be carried out in the next stages of qualification: DQ, IQ, OQ, PQ.

Design Qualification - The DQ protocol is the documented evidence that the equipment purchased has the appropriate characteristics, in accordance with the requirements and specifications defined in the URS. The tests defined are:
- Record of available documentation
- Compliance with regulatory requirements
- Compliance with user requirements
The results of DQ allow to conclude that there is documented evidence that all tests included in the DQ protocol were performed correctly with the final result in accordance with the requirements, so the equipment is qualified in terms of design.

Installation Qualification - The purpose of the IQ is to establish documented evidence that the portable Raman is properly installed to specifications. The tests carried out for the qualification of the installation were:
- Technical documentation of the equipment
- Characterization of the equipment and its main components
- Configuration of control devices
The results of IQ conclude that there is documented evidence that all tests included in the IQ protocol were performed correctly with the final result in accordance with the requirements, so the equipment is qualified in terms of installation.

Operational Qualification - The OQ serves to establish documented evidence that the equipment operates properly according to specifications. The tests performed were: SOP (Standard Operational Procedure), Basic Operational Functions, Alarms and malfunctions, Management and Control Parameters, Password Access Levels, Report management, Training of Operators, Wavelength Accuracy Test, Spectral Resolution Test, Testing Technical Functionality.

The OQ results were all positive and allow to conclude that there is documented evidence that all tests included in the OQ protocol were performed correctly with the final result in accordance with the requirements, so the equipment is qualified in terms of operationality.

Performance Qualification – While OQ tests are performed annually through OVP software (OPUS Validation Program), the PQ is performed daily. The tests for PQ are:
- Wavelength accuracy test
- Wavelength precision test
- Photometric accuracy test
- Laser power test
Additionally, the efficacy of the sample verification test was tested, by changing the name of samples to verify if the system analyzes correctly the sample inserted on the equipment. The PQ results allowed to conclude that there is documented evidence that all tests included in the PQ protocol were performed correctly with the final result in accordance with the requirements, so the equipment is qualified in terms of performance.

Acceptance Final Report
The qualification process is finalized by elaborating a document that compiles all the data from the previous reports defined initially in the master plan. This document describes the deviations that have elapsed throughout the process and a conclusion/evaluation is presented. If all stages of the qualification are conformed, then it can be concluded that the equipment is suitably qualified.

For each of the previous qualification steps is indicated that tests were properly executed with the result conforms the requirements. In addition, this document also serves to compile possible observations on corrective actions and whether or not deviation reports have been issued. In the case of this particular equipment, all stages were successfully completed, no deviations were recorded, and the results were compliant with the defined requirements. In this way, it is possible to affirm that the portable Raman equipment, object of study of this work, meets the qualified state and is suitable for the intended use which is RM verification.

4. Results of Database: Development and Validation
The Raman spectrometer allows the verification of a sample by its direct comparison against the spectrum of the reference sample included on the library. Thus, for each RM analyzed, it is necessary to have the corresponding reference sample spectrum. The set of all reference spectra constitutes the data library. This library has to be created individually for each sample and it is necessary to enter the following fields: Material Name; Compound Name, Code, Lot and Container.

The key to the successful development of the data library and the use of Raman is the "Compound Name" field, which is the way to chemically group species but with different names. This concerns the fact that there are the same distinguished RM with different material codes and names because they come from different suppliers. Thus, a previous work had to be done, where a code field had to be associate with the CAS number (Chemical Abstracts Service number). This CAS number is the specific universal designation of a chemical.

Reference Samples
RM samples were previously identified by the methods of the European pharmacopeia already validated in the laboratory. Samples of solid RM were packed in polyethylene bags, while the liquid samples were collected into lyophilized glass vials, all carefully identified. In order to obtain the best spectra quality, various types of packaging and their influence on the verification result were tested.

Packaging
The thickness of the bags used for collecting samples in the LAB (laboratory) and the bags received in the warehouse is totally different. To study the impact this could have on the analysis, several measurements were simulated in the LAB with successively increasing thickness bags. Analyzing the data
obtained in this test allowed to conclude that even with a 12 folded plastic bag it is possible to have an interesting result: HQ=0.9816, which means that the sample is 98.16% similarly to the reference spectrum. The clear glass vials and amber glass vials were also compared against each other and against the plastic bag. It is possible to conclude that the best storage conditions correspond to the transparent plastic bag, followed by the clear glass vials and finally the amber glass vials.

**Methods Development**

A general method has been developed that applies to most RM, following the collecting procedure previous described. Whenever RM is received with new identity codes it is necessary to make an analysis to see if they can be included in this general method or if it is a case that requires more attention to be inserted in one of the individual methods. At the moment, the general method has 289 RMs, which corresponds to more than 50% of the total RMs.

**Individual Methods**

There are raw materials that supplied by different manufacturers and different batches and physical variations justify the development of an isolated verification method capable of introducing maximum variability. For this, an individual analysis of each reference spectrum was performed, which resulted in the creation of the 7 following methods: Paracetamol; Titanium dioxide; Crosycarmellose Sodium; Stearic acid; Cetostearyl alcohol; Hydrogenated Ricinium Oil; Vegetable oil.

The need to develop specific verification methods for a RM group was triggered by the intention to use the equipment for the verification not only of isolated RM, but of compounds.

**Methods for Mixtures**

As example, Sepifilm is an excipient composed by (Figure 3):

- Titanium dioxide (marked green)
- Stearic Acid (marked in red)
- A cellulosic fraction (marked in dark blue).

The peaks recorded in Sepifilm correspond to the sum of the peaks of stearic acid, cellulose and titanium dioxide. The differences between the individual compounds and the mixture are negligible, considering the percentage variation of the peaks marked in the individual materials when compared to the composite Sepifilm, represented on the Table 2:

<p>| Table 2 – Peaks recorded by Sepifilm, Stearic Acid, Titanium Dioxide, Cellulose. |
|------------------------------------------|-------------------------------|-------------------------------|-------------------------------|</p>
<table>
<thead>
<tr>
<th>Sepifilm (cm(^{-1}))</th>
<th>Stearic Acid (cm(^{-1}))</th>
<th>Titanium Dioxide (cm(^{-1}))</th>
<th>Cellulose (cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>395,09</td>
<td>394,74</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>514,80</td>
<td>514,01</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>637,31</td>
<td>637,24</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>890,82</td>
<td>892,09</td>
<td>-</td>
<td>891,57</td>
</tr>
<tr>
<td>1127,74</td>
<td>1128,22</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1295,36</td>
<td>1295,25</td>
<td>-</td>
<td>1292,74</td>
</tr>
<tr>
<td>1370,08</td>
<td>-</td>
<td>-</td>
<td>1377,91</td>
</tr>
</tbody>
</table>

The same analysis was done for Omeprazole pellets and Lanzoprazol pellets with similar positive results.

**Database Validation**

Given the importance of the library in the successful use of Raman spectra to identify RM, before the equipment came into use, the developed library was validated. For the creation of the library it was selected as criterion to create library only for the RM received by Sofarimex between 2014 and 2017.
To validate the library, we selected some criteria that introduced some variability namely:

- Different batches
- Different suppliers;
- Different granulometry;
- Different physical states;
- API vs Excipient.

Once the reference library is developed, it is important to validate it to ensure its performance and reliability. The validation of spectral libraries for purposes of identity verification is usually performed by determining their false negative and false positive rates.

**OVP Validation**

With this functionality the desired library is selected, and the software correlates all the spectra to each other, in order to assign an HQI between the most similar correspondences. Each standard spectrum corresponds to one entry in the library. If there is HQI > 95% between 2 different compounds, the library is not validated and the conflicting compounds are marked in red.

All methods created have been validated and are suitable for use in the sampling raw material.

**Internal Validation**

In addition to OVP validation, an internal validation was made, in order to analyze the critical factors previously discussed. Thus, 14 LAB samples were selected, of which 6 were API, 4 were excipients and the other 4 were external Bulk.

**Repeatability**

One of the objectives of this validation was to obtain conclusions regarding the reproducibility of this analytical method. To test the repeatability of the readings, 3 replicates were performed for the 14 RM codes. Analyzing the results of the corresponding HQI it is safe to say that the equipment reveals a notorious repeatability, which is a fundamental factor to ensure the serial use in the warehouse.

**Different batches**

Production differences can occur between batches. Therefore, it was considered pertinent to include in the library validation 2 different batches.

---

**Figure 3 - Sepitlm spectrum (yellow), Titanium Dioxide (green), Stearic Acid (red), Cellulose (dark blue).**
batches of each RM. The results of this study were conclusive indicating the absence of influence on the results of the different batches. This fact values the flexibility of the library, ensuring that when different batches are sampled from those used to construct the library, the results are unequivocal, independent of this factor.

**Different Manufacturers**

Another point of discussion is whether a library should be vendor specific or not. It is natural that some RM vary considerably from supplier to supplier, in terms of quality and purity, particle size, humidity, etc. These differences may result in spectral differences, which may cause the threshold to decrease and thus the degree of trust of the library. To prevent this from happening, it is recommended to create a vendor-specific library, for example, a library of material X for vendor A and vendor B. For this reason the database was created with a single spectrum for each code, and RM from different suppliers originate distinct internal codes.

**Chemically Similar RM**

Samples with a high degree of similarity at the molecular level may be a limitation for the use of this technology in the verification of PM. The goal is to be able to distinguish all compounds, even if similar. As an example of this type of compounds arise the celluloses and their variants. Given the importance of distinguishing them, the samples belonging to this molecular group were selected: Hydroxypropylmethylcellulose; Ethylcellulose; Hydroxyethylcellulose; and Hydroxypropylcellulose. Although they only differ in a chemical group, the different molecular bonds translate their differences in the correspondent peaks.

**5. Impacts of Raman Implementation**

The sampling plans dictate how the sampler operators will harvest the samples of the RM to be collected. Thus, where Raman can be used, operators use the equipment to check all containers in the batch, but only collect samples of $\sqrt{n} + 1$ containers, where $n$ is the number of containers that make up a batch, according to WHO (World Health Organization) sampling guideline.

It was also necessary to change the operational procedures related to sampling, where the sampling plans are described and according to which the samplers and all those involved in the process are ruled.

Finally, it was considered important to define a contingency plan to include in the procedures, to guarantee that all containers are properly identified. For this, we simulated what could go wrong in the use of Raman and defined an alternative procedure. In the case of a wrong identification of a container via Raman, it is considered that this negative result should not be accepted without repeating an analysis. Thus, if in a batch with multiple containers all are correctly identified except one, what one should do is to take a sample of that container in an individual flask and segregate it to be identified in the LAB by methods described on the European Pharmacopeia. Because the factors that may influence the Raman analysis are known, the identification result of the LAB will prevail over that of the Raman.

**6. Conclusions and Future Perspective**

It is evident that this technology has many advantages, such as: reduction of analysis in the LAB, reduction on time analysis, cost reduction, more savings on resources used,
more availability of analysts to conduct further tests.

One of the future objectives is to increase sampler autonomy at maintaining and developing the library databases. At the time of implementation of the equipment, the general method consisted of 289 RM, corresponding to more than 60% of the total RM number. There are still several aspects to improve, namely in the use of the equipment for the verification of liquid RM and for solid RM packed in paper bags, making it impossible to use this equipment. Another limitation is related to heterogeneous RM, with different granulometry. Some bugs of the OPUS software and OVP functionality were also observed, which could be easily solved with a future update of the software.

Considering the ease of use of this technology and the speed of analysis, a future suggestion is the acquisition of another equipment to cover the two existing sampling rooms. In addition, another interesting application for Sofarimex would be the adoption of this technology in LAB to identify the finished product. The ability to identify RM through blisters was tested and the results were positive. The effectiveness in identifying the packaging material, namely the various types of plastic received in reels, has not been tested.

Despite the difficulties experienced during this work, the success of the implementation of this technology is with main consequences in the reduction of the number of tests to be carried out in LAB, the decrease of the probability of cross contamination, as well as the reduction of the sampling material resources.

7. References


