

Detection of low numbers of bacterial cells in pharmaceutical drug product using Raman Spectroscopy and PLS-DA multivariate analysis.

Grosso R.A.^{1,2}, Walther A.R.¹, Brunbech E.², Sørensen A.², Schebye B.³, Olsen K.E.², Qu H.¹, Hedegaard M.A.B.¹, Arnspang E. C.¹.

¹ Department of Green Technology, SDU- Biotechnology, University of Southern Denmark, Odense, Denmark.

² Product Supply Injectable Finished Products, Microbial Competence Centre, Novo Nordisk A/S, Copenhagen, Denmark.

³ Product Supply Injectable Finished Products, Technology Innovation, Novo Nordisk A/S, Copenhagen, Denmark.

Abstract

Sterility testing is a laborious and slow process to detect contaminants present in drug products. Raman spectroscopy is a promising label-free tool to detect microorganisms and thus gaining relevance as future alternative culture-free method for sterility testing in pharmaceutical industry. However, reaching detection limits similar to standard procedures while keeping a high accuracy remains challenging, due to weak bacterial Raman signal. In this work, we show a new non-invasive approach focusing on detect different bacteria in concentrations below 100 CFU/ml within drug product containers using Raman spectroscopy and multivariate data analysis. Even though Raman spectra from drug product with and without bacteria are similar, a partial least squared discriminant analysis (PLS-DA) model shows great performance to distinguish samples with bacteria contaminants in limits below 10 CFU/ml. We use spiked samples with bacteria spores for independent validation achieving a detection accuracy of 99%. Our results indicate a great potential of this rapid, and cost-effective approach to be use in quality control of pharmaceutical industry.

Introduction

Testing for microbial contamination is a crucial step in quality control of pharmaceutical drug products (DP) before their commercial release ^{1,2}. Standard procedures for bioburden testing are highly time-consuming, costly and are limited in terms of sensitivity and specificity given that they depend on growing conditions. ^{3,4}. Hence, the pharma industry needs to develop and implement faster and cost-effective novel technology for this purpose. Bacterial contamination criteria vary depending on the type of pharmaceutical formulations and administration route ^{5,6}. The complete absence of micro-organisms is required for quality control tests of drug products required to be sterile ⁷. For this reason, novel biotechnological approaches for contamination tests must secure detection of all microorganisms potentially present in the final product ^{2,8,9}. Analyzing for absence of all micro-organisms by traditional growth based microbiological methods presents a further challenge for the laboratory environment to minimize the risk of false positive test results caused by laboratory contaminations during sample handling ⁶. Within the last decades new set ups involving spectroscopic technology have been studied for faster and simpler microbial detection and quantification in drug products ⁹⁻¹².

32 Raman spectroscopy (RS) is a non-invasive method based on inelastic scattering of monochromatic light upon
33 interaction with chemical bonds present in the sample. Each molecule gives a specific Raman spectrum (fingerprint)
34 depending on their chemical environment, and chemical and biophysical properties¹³. In combination with
35 chemometrics, RS has been gaining significant consideration in the pharmaceutical industry given its reduced cost,
36 faster quantitative analysis, and real-time monitoring of various processes that involve changes of the molecules
37 and/or their chemical environments^{14, 15}. Concerning microbial contamination, RS has shown great performance
38 detecting a wide range of bacterial components i.e., lipids, proteins, amino acids, nucleic acids¹⁶. This technique is
39 even able to distinguish among several bacterial strains present in water-based formulations and in solid drugs with
40 high accuracy^{17, 18}. However, studies with RS showing high robustness while retaining the ability to distinguish even
41 bacterial strains lack sensitivity and do not reach a sufficiently low limit of detection (LOD) of bacteria cell number
42 for bioburden test acceptance^{19, 20}. Contrary, more sophisticated RS set ups with LOD below 10³ CFU/ml have
43 demonstrated drawbacks in terms of reproducibility, data analysis and expensive technology making these methods
44 inconsistent and difficult to scale up for pharmaceutical industry application^{21, 22}.

45 Even though advances with RS are promising, new approaches must improve its accuracy and reproducibility to be
46 suitable for pharmaceutical industry quality control, and act as simpler, faster, and more cost-effective applications.
47 Remanent challenges when detecting bacteria with RS in pharma products are (i) discrimination between Raman
48 spectra from organic molecules present in the formula and bacterial ones, (ii) detection at low contamination given
49 the weak signal from the bacteria in comparison with the product volume (iii) contribution to the Raman signal
50 from other sources such as product packaging, fluorescent compounds, and (iv) correct data processing and
51 statistical analysis model¹⁴. In this study, we present a novel approach to up-concentrate and detect ≤10 CFU/ml
52 of relevant bacteria with RS and multivariate analysis without breaching the primary DP package. The outcomes of
53 this project support RS as non-invasive and non-destructive method to detect bacterial contamination in DP as
54 alternative to the pharmacopeial⁷ destructive method susceptible to laboratory contaminations. Furthermore, the
55 developed method has the potential to enable real-time monitoring of the contaminations in pharmaceutical
56 processing.

57 **Experimental**

58 **Reagents and bacteria strain.**

59 The pharmaceutical product (DP) contains 3 ml of solution with 100 U/ml insulin, glycerol, 1.80 mg/ml phenol (CAS
60 108-95-2), 2.06 mg/ml m-cresol (CAS 108-39-4) (preservative), zinc acetate, glycerol, phosphate buffer, sodium
61 chloride and water for injections. The primary package is a vial made of 2 mm borosilicate glass.

62 Freeze-dried preparations in BioBall® MultiShot 550 of *B. subtilis* (NCTC 10400) and *S. enterica* (ACM 5080) were
63 from bioMérieux. *B. subtilis* in spore form. *S. haemolyticus* (ATCC 29970) was kindly provided from Microbial
64 Competence Centre, Novo Nordisk. Other reagents were Ethanol (20C184005, VRW chemicals), Phosphate buffered

65 saline solution (S3308, MP Biomedicals), Luria-Bertani broth agar (SLCC1516, Sigma) and Tryptic soy broth (TSB)
66 (ICNA091010717, MP Biomedicals). Tryptic soy agar (TSA) was prepared by adding 20 g/L of agar (Sigma) to TSB
67 solution.

68 **Bacterial culture and growth conditions.**

69 For experiments with vegetative forms of *B. subtilis*, one BioBall® MultiShot 550 of *B. subtilis* was inoculated in LB
70 broth media and incubated at 35 °C for 18-24 hour and then sub-cultured in LB agar and incubated in aerobic
71 conditions at 30-35 °C for 24 h. Bacterial numbers were estimated by colony harvesting, dilutions in sterile PBS and
72 counting in chamber under the microscope, then applying dilution factor calculation. Experiments with low CFU
73 number (50 and 10 CFU) were performed directly with dilutions from the dissolved BioBalls (550 CFU/pellet). *S.*
74 *enterica* and *S. haemolyticus* were cultured using TSA plates. Experiments with low numbers of *B. subtilis* spores
75 were made by dilutions from BioBall® MultiShot 550 of this bacterium and then spiking the samples with the
76 respective amount to reach a final concentration of 50 and 10 CFU/ml. Controls of injected CFU number were made
77 by enumeration in TSA plates and direct counting of colonies.

78 **Sample preparation and Raman spectroscopy**

79 Spiked vials for experiments with 3×10^8 , 50 and ≤ 10 CFU/ml of bacteria were prepared using the following
80 procedure. Briefly, in a laminar air flow (Heraguard ECO 0.9, Thermo Scientific), bacterial dilution was prepared
81 using sterile DP solution as diluent and homogenized by vortexing. Then the precise volume was injected into the
82 products vials (samples) using a calibrated syringe (Hamilton® syringe 1700 series) to get the desired CFU
83 concentration. The vials were placed in a self-built plate and centrifuged at 2500 rpm for 17 minutes. Before
84 experiments, vials' outer glass surface was cleaned with ethanol avoiding any Raman signal from organic material
85 adsorbed during sample handling. Given that the DP contains preservatives, a parallel control was run covering the
86 whole procedure time from CFU injection till last sample measurement with RS. This control was cultured along
87 with analyzed sample to evaluate bacteria survival without exposure to laser.

88 Raman spectra were recorded at room temperature using a multimode diode laser (FATBOY model, Ocean Insight)
89 emitting at a wavelength of 785 nm. The laser excitation was fiber-coupled to a focusing 785 nm immersion Raman
90 probe (Ocean Insight) with a 5 mm focal length delivering a total laser power of 350 mW at the sample. The Raman
91 back-scattered light collected by the probe was directed through a low OH multimode fiber to the spectrometer
92 (QE Pro Raman Series, Ocean Insight) equipped with a charge coupled device operated at -50 °C. For each sample,
93 10 Raman spectra were acquired with an integration time of 10 seconds. Spectrometer setup, data acquisition and
94 control were done with Ocean View 2.0.8 software (Ocean Insight).

95 **Data processing and analysis.**

96 Raman spectra were pre-processed using MATLAB R2021a software (MathWorks, US). Data was truncated to the
97 wavenumber region 400-3200 cm^{-1} , baseline corrected using asymmetric least squared smoothing²³, normalized
98 (area under curve) and mean centered. PLS-DA²⁴ was performed using the PLS-regression algorithm in MATLAB.
99 For model calibration, we used a total of 100 spectra of sterile and 100 spectra of vegetative *B. Subtilis*. For
100 experiments with *S. enterica*, *S. haemolyticus*, and *B. subtilis* spores a total of 80 Raman spectra from sterile and
101 spiked samples were analyzed. To differentiate *B. Subtilis*, *S. haemolyticus* and *S. enterica* from sterile conditions a
102 10-fold cross validated PLS-DA model was applied to each case with 50 and 10 CFU/ml. For *B. Subtilis* spores, *S.*
103 *haemolyticus* and *S. enterica* four, three and three components were used in the model respectively based on the
104 results of the cross validation (**see Table 1**). An independent validation set of spectra from *B. Subtilis* spores was
105 tested against the corresponding model calibration to obtain the specificity and sensitivity of the model.

106 Results and discussion

107 Proof-of-concept

108 We aimed to generate a novel approach using dispersive RS associated with partial least squares discriminant
109 analysis (PLS-DA) to detect bacteria within the DP primary packaging. The first goal was to analyze if the Raman
110 spectra from the bacteria could be distinguished from those related to molecules present in the DP. A *proof-of-*
111 *concept* was designed to up-concentrate bacteria in a smaller area of the DP container and thus increasing the
112 chance for the laser to find the contaminants. For this purpose, the samples were centrifuged in special plates
113 holding the vials in an inclined upside-down position. This method was efficient in localizing the bacteria along the
114 upper part of one side of the product vial, i.e., close to the neck. To obtain a clearer bacterial spectrum for model
115 training vegetative form of *B. subtilis* were injected into five DP samples with a final concentration of 3×10^8 CFU/ml.
116 Raman spectra from spiked samples with vegetative *B. subtilis* (SS-vBS) and non-injected negative controls (NC)
117 were taken covering the whole area within the vials where the bacterial pellet was localized after centrifugation.
118 Raman spectra after baseline correction and normalization of each group are presented in **Fig. 1A**. Both groups
119 presented similar spectra but with small variations in intensity in peaks related to bacterial-associated organic
120 molecules observed in the wavenumber ranges 700-1800 and 2800-3200 cm^{-1} (**Fig. 1A**). PLS-DA model was selected
121 due to its power to maximize the inter-class variation and its prediction capacity among samples with unknown
122 degree of within-group variability^{17,24,25}. Considering up-till ten components, analysis of variance explanation (**Fig.**
123 **1B**), indicated that three components represented 86.4% of the variation on the dataset. A smaller variation was
124 enclosed in components four and five. Afterwards, any extra component added did not solely explain a significant
125 part of the remaining variation (**Fig. 1B**). PLS-DA classification plot (**Fig. 1c**) showed that a three-component model
126 was sufficient to discriminate the Raman spectra from SS-vBs and NC. Variation observed among samples, mainly
127 on SS-vBS, could be related to measurements from different locations from these vials. Closer inspection of the
128 loading vectors (LV) showed distinct bacteria-associated peaks, which weighted positively for SS-vBS on the score
129 plot (**Fig. 1d-e**). SS-vBs received positive scores for peaks on the wavenumbers 786, 814-850 and 1090 cm^{-1} (nucleic

130 acid region), 1000 cm^{-1} (aromatic amino acids), 1060 -1200 cm^{-1} (fatty acids C—C stretching), 1440 cm^{-1} (C—H
 131 vibration), and 1656 cm^{-1} (amide I vibration) among the most relevant (**Fig. 1d-e**)^{19, 22, 26, 27}. Some bands located in
 132 regions 703-760 and 820-840 cm^{-1} indicating aromatic amino acids can also be related to preservatives (m-cresol
 133 and phenol), present in the product formula^{28, 29}. The high wavenumber region 2800-3200 cm^{-1} (C—H stretching
 134 region) was kept in the analysis since it has previously been related to major bacterial macromolecules^{26, 30} that are
 135 not present in the DP. LV1 and LV2 presented peaks with negative intensity amplitude in the Raman shift 746, 1000,
 136 1029, 1090, 1 cm^{-1} (**Fig. 1e**). The latter were important in the NC samples, which received negative scores specially
 137 in LV1 (**Fig. 1d-e**) possibly related to the absence of bacterial structures. The negative bump observed in the region
 138 at 1300-1450 cm^{-1} in LV1 (**Fig. 1e**) is mainly associated to BO_4 and SiO_4 bonds from the primary packaging made of
 139 borosilicate glass³¹.

140 A 10-fold cross-validation was performed to analyze the number of components needed in order to obtain the best
 141 performance of classification through the root mean squared error of cross-validation (RMSECV) and comparing
 142 with the overall fitting by RMSE of calibration (RMSEC) (**Fig. 1f**). Based on RMSECV (**Fig. 1f – orange line**), five
 143 components were the limit for error reduction and onwards we had a diminishing in return. However, three
 144 components were enough to perform without misclassification reducing the risk of overfitting and making the
 145 model more robust in terms of classifying future samples that are different to the training dataset. In the **Table 1**
 146 are summarized all RMSEC and RMSECV values obtained for each experiment as well as their respective number of
 147 components considered, and the outcome of samples misclassified.

Table 1: Values of root-mean squared error of calibration (RMSEC) and cross-validation (RMSECV) of each experimental model considering level of detection of each bacterium and the number of components included. Misclassified samples outcome after PLS-DA analysis.

Experiment	Detection level (CFU/ml)	Number of components	RMSEC	RMSECV	Misclassified samples
<i>B. subtilis</i> (V)	3×10^8	3	0.28	0.31	None
<i>S. enterica</i>	50	3	0.058	0.066	None
	≤ 10	3	0.062	0.073	None
<i>S. haemolyticus</i>	50	3	0.18	0.2	None
	≤ 10	3	0.29	0.33	None
<i>B. subtilis</i> (S)	50	4	0.29	0.45	None
	≤ 10	4	0.28	0.63	None

148

149 In our system we used 785 nm visible light, since water has a relatively low absorption at this wavelength while
 150 obtaining a relatively strong signal from bacteria in comparison to other wavelengths¹⁴. In addition, the 785 nm

151 laser presented less fluorescence interference and improved signal-to-noise ratio compared to other visible light
152 spectra, such as a 532 nm laser³²⁻³⁴. Supporting this choice, previous studies have shown great robustness of 785
153 nm in detecting wide spectrum of both Gram-positive and Gram-negative bacteria³⁵. In summary, this concept
154 adding a special centrifugation of vials could improve the bacterial detection with RS in DP. Moreover, the PLS-DA
155 approach allowed us to maximize the inter-class differences and therefore discriminate the Raman signal of bacteria
156 suspended in DP from the DP alone within the primary packaging.

157 **Detecting low number of vegetative bacteria within primary package.**

158 Major challenges encountered when reaching low CFU detection limits with RS are: (i) to discriminate bacterial
159 signal from other organic compounds present in the DP without compromising the model accuracy, (ii) to detect
160 bacteria in concentrations lower than 100 CFU/ml keeping the cells viability for future identification procedures,
161 and (iii) to detect bacteria and discard the primary packaging signal while keeping the integrity of the DP primary
162 container^{14, 36-38}. Therefore, we challenged our RS-PLS-DA model to detect vials containing below 100 CFU/ml of
163 two important yet different bacteria: *Salmonella enterica* (rod-shaped gram-negative) and *Staphylococcus*
164 *haemolyticus* (coconut-shaped gram-positive). It is important to mention that we did not focus on distinguishing
165 between the bacterial species but on confirming the presence or absence of contaminants with high reproducibility.

166 Following the same procedure of spiking and centrifugating as explained before, we prepared samples with each
167 species reaching a final concentration of 50 CFU/ml and ≤ 10 CFU/ml. Raman spectra were obtained from SS
168 containing 50 CFU *S. enterica* (50-SE), ≤ 10 CFU *S. enterica* (10-SE), 50 CFU *S. haemolyticus* (50-SH) and ≤ 10 CFU *S.*
169 *haemolyticus* (10-SH) and compared against NC (**Fig. 2**). In SS, given the inclined centrifugation of the vials we
170 expected to obtain a mixture of positive bacterial signal from the first half of the vial and negative onwards. After
171 Raman analysis the samples were cultured to compare results with the ones obtained from RS. Explorative analysis
172 of the training dataset was performed using the same PLS-DA model and corroborate if consistent output variables
173 were obtained i.e., discrimination between samples with and without bacteria. PLS-DA classification and score plots
174 of loading vectors are shown in **Fig. 2a-h**. A three-component model showed a well separation of samples with
175 bacteria and NC of both concentrations of *S. enterica* and *S. haemolyticus* (**Fig. 2a,c,e,g**). Both experiments with *S.*
176 *enterica*, 50-SE and 10-SE, received a better discrimination in terms of inter-class distance (Y-axis) from NC group
177 in comparison to *S. haemolyticus* samples (**Fig. 2a, c**). However, score plots of LV1 and LV2 showed a correct
178 discrimination of SS and NC training dataset clusters even at two-component model for all bacterial specie and
179 concentration analyzed (**Fig. 2 b,d,f,h**). On the score plots, groups representing SS with 50-SE and 10-SE appeared
180 as two separated subclusters with different weight scores for LV1 and LV2 (**Fig. 2b,d**). This intra-group gap reflected
181 the measurements from areas with bacterial signal and without bacteria from centrifuged samples bearing bacteria.
182 Concerning *S. haemolyticus*, in both 50-SH and 10-SH groups the inter-class separation was less prominent
183 compared to *Salmonella* and observations from the negative areas appeared closer to the sterile group (**Fig. 2e,g**).
184 This was also demonstrated in the score plot where three sub-clusters were defined for SS with 50-SH and 10-SH,

185 indicating more variability among these observations (**Fig. 2f,h**). In these plots it is also observed that some
186 observations from SS at both concentrations scored similar to sterile groups. Despite that, no misclassification of
187 training dataset was obtained (**Table 1**). The difference observed in term of greater discrimination for SS with *S.*
188 *enterica* in comparison to *S. haemolyticus* could be related to bacterial composition, size and shape being reflected
189 in Raman spectra peaks intensity³⁵. Supporting this idea, experiments detecting both 50-SE and 10-SE the RMSECV
190 obtained was remarkable lower under three-components compared to detection of 50-SH and 10-SH (**Table 1**).
191 Bacteria culture was representative of Raman results obtaining growth of bacteria colonies for all SS evaluated and
192 no growth on the NC (data not shown)

193 **Low LOD of bacterial spore detection**

194 Some bacteria relevant for pharmaceutical production have the ability to produce spores as a survival mechanism
195 in a hostile environment. A method used for quality control of pharmaceutical presentations required to be sterile
196 should also be able to detect spores, especially given the increased likelihood of bacteria in spore form surviving
197 unfavorable conditions for extended time periods and reaching the patient. Hence, we challenged our RS-PLS-DA
198 model to detect low concentrations of *B. subtilis* spores (BSs). We spiked vials with 50 CFU and ≤ 10 CFU/vial of *B.*
199 *subtilis* spores (50-BSs and 10-BSs respectively) and then analyzed with RS and compared to NC. In **Figures 3a** and
200 **3c** is presented the PLS-DA classification for both 50-BSs and 10-BSs respectively. The model correctly discriminated
201 the samples even though in both concentrations the variation between classes was less prominent in comparison
202 to the case with vegetative cells (**Fig. 2**). The less prominent discrimination was supported by the intra-sample
203 variation as well as the limited inter-class separation observed on the Y-axis (**Fig. 3a, c**).

204 For spore detection at very low cell numbers, four components were needed to reach a complete group separation
205 instead of the two-component needed for vegetative cells. This was observed on the score plots were overlapping
206 among the observation was remanent when comparing LV2 vs LV3 (**Fig 3b, d**). In addition, in the RMSECV plot an
207 evident improving step in error reduction could be seen when jumping from three components to four (**Fig. 3e**). In
208 a general perspective it is clearly observed that RMSEC and RMSECV obtained in both spore concentrations were
209 higher compared to the errors observed in vegetative bacterial forms (**Table 1**). Loading plot of LV1-4 showed weak
210 but clear Raman signal related to spores which was mainly supported by the peaks at 786, 814-850 and 1090 cm^{-1}
211 (nucleic acid), 1464 (lipids), 1579, 1665 cm^{-1} (proteins) and 2800-3200 cm^{-1} (C—H)^{30, 39, 40} (**Fig. 3f**). At this spore
212 concentration, previously reported specific peaks related to *Bacillus sp.* spores in the wavenumbers 824, 1017 1395
213 and 1579 cm^{-1} ^{30, 39, 41, 42} were difficult to spot given their weak signal in comparison to other peaks. (**Fig. 3f**). The
214 latter is possibly related to different factors, first that the model was built and trained based on vegetative forms
215 which are phenotypically different to spore form^{19, 26}. The second reason could be associated to the low working
216 concentration making it more difficult to visualize spore-specific peaks in comparison to larger and more
217 predominant macromolecules, which already had a weak signal in the LVs. Spores are around one-third of the size
218 of a vegetative cell^{43, 44}, so it is more difficult to obtain a clear Raman signal from them and thus also to detect low

219 numbers of spores. Is important to highlight that even if most authors consider wavenumbers from 400 to 1800
220 cm^{-1} as the most active region for bacterial Raman signal, with the spores we should also consider the high
221 wavenumber region 2800-3000 cm^{-1} . The latter region is related to predominant spore components such as fatty
222 acids, carbohydrates, and proteins, which are packed together with the genetic material ^{19, 26}. These results
223 together demonstrate that our RS-PLS-DA model is not affected by bacterial phenotypic features but instead
224 focuses on common components of them.

225 **Model validation**

226 Cross-validation of independent samples is crucial to demonstrate the ability of the model to discriminate new
227 samples with unknown contamination. For this purpose, we performed a 10-fold leave-one-out cross-validation
228 analysis using our RS-PLS-DA model and a new independent set of samples. Contamination with *B. subtilis* spores
229 were used given they were the most challenging case concerning detection of low numbers. Two datasets were
230 used for cross-validation: (i) a training data set using samples from **Fig. 3c** and (ii) an independent (validation) set
231 with 20 new samples comprising ten vials with ≤ 10 CFU-BSs and ten sterile vials. After Raman analysis, samples
232 were cultured to compare results with the ones obtained by RS-PLS-DA. Classification of training (stars) and
233 validation (triangles) datasets from cross-validated data are presented in **Fig. 4a**. Both sterile (blue triangles) and
234 contaminated (red triangles) validation groups were discriminated with 99% specificity and 98.3% sensitivity,
235 establishing a hard decision cut-off in zero (dotted line) (**Fig 4a**). Validation samples dataset presented larger intra-
236 group variations compared to the training group. This represented a realistic scenario of new samples being
237 compared with the training dataset building the PLS-DA model. RMSE of validation groups (RMSEV) (**Fig. 4b, blue**
238 **line**) showed that an improvement could be obtained up till five components considered. This reduction was not
239 seen by analyzing RMSECV of the independent validation (black line) where after the third component any
240 additional component did not reduce the prediction error (**Fig 4b**). The latter is important when we take into count
241 that the results showed in **Fig. 3**, indicated that four components were needed in order to discriminate 50 and 10
242 CFU of BSs from NC without misclassification (**Table 1**). Therefore, using less components some samples crossed
243 the decision line generating some false positive and false negative observation which impacted on the accuracy
244 obtained (**Fig 4a**).

245 In this work, we obtained similar values in terms of accuracy in comparison to previous reports showing sensitivities
246 and specificities around 95-99% using RS coupled with multivariate analysis models. However, some approaches
247 do not specify the limit of detection reached ⁴⁵ or they presented theoretical detection limits below 100 CFU/ml
248 based on RMSECV values obtained on the calibration and not in physical measurements ⁹. Another common
249 drawback is to ensure reproducibility of the results using a single model to detect samples with less than 10 CFU/ml
250 or when cross-comparing among different species ^{22, 26, 35, 37}. In addition, our results showed a great performance
251 about the grade of robustness including three highly different bacteria species and bacterial spores. Another
252 important feature is that we used a non-invasive method to reach low LOD which allowed us to test the DP within

its container. The latter is an improvement of previous concepts involving complex RS systems to obtain low LOD involving procedures breaching the integrity of the DP samples such as needing external amplifier particles and/or dried samples on metallic surfaces^{37, 46}. In this sense, sample handling that requires breaching of the barrier presented by the original container also poses a risk of obtaining secondary sample contaminations thereby giving a false positive result. Time consumption is also a drawback when analyzing samples with RS concepts involving several steps, especially if the final application is in pharmaceutical industry. Even though LOD below 10³ CFU/ml could be secured, these kinds of multiple-step and advance systems need many hours to analyze a single small sample³⁷ On the contrary, we only needed around 3 minutes to evaluate each sample containing 3 ml of product with RS after centrifugation, and the evaluation is done without breaching the barrier of the original sample container. Further analyses are needed to deepen in understanding the limits of RS technique for application in contamination control such as detection of cell debris, death cells and other relevant scenarios that may cause ambiguous results.

Conclusions

In the present study we evaluated a fast and non-invasive approach to discriminate pharmaceutical products vials containing low numbers of bacteria from sterile ones using dispersive Raman spectroscopy in association with PLS-DA. The RS-PLS-DA concept was challenged to detect bacterial species suspended in drug product within its primary package. Three highly different bacteria including *B. subtilis* spores were detected without breaching the DP vial. This rapid and simple concept innovates in its effective way to localize the contaminants in a smaller area within the product vial in order to secure laser detection. Our results showed a successful discrimination when detecting vegetative cells and spores at the very low concentration of 10 CFU/ml, even in the presence of other organic molecules from the product formula in the intact DP primary container. True independent validation showed an outstanding performance with high sensitivity and specificity. In summary, we provide a feasible approach using RS in association with PLS-DA to detect extremely low numbers of cells or spores with high accuracy and reproducibility without compromising the robustness of the method. These results support Raman spectroscopy as a promising biotechnological tool suitable for bioburden test in quality control of pharmaceutical industry.

Author contributions

G.R.A. conceived and performed the experiments, analyzed the data, discussed the results, and wrote the manuscript. W.A.R. and H.M.A.B. performed the data processing script and analyzed the data. S.B. provided project design ideas and discussed the results. B.E and S.A provided expertise for experimental design, discussed the results, and provided feedback. O.K.E and Q.H. provided expertise and feedback. A.E.C. coordinated the project and provided expertise, feedback. All authors read and approved the final version of the manuscript.

Acknowledgment

285 Villum young investigator given to A.E.C (19105). Novo Nordisk A/S to A.E.C., for financial support for reagents
286 acquisition.

287 References

- 288 1. WHO, 2003.
- 289 2. M. Jiang, K. A. Severson, J. C. Love, H. Madden, P. Swann, L. Zang and R. D. Braatz, *Biotechnol Bioeng*, 2017,
290 **114**, 2445-2456.
- 291 3. M. R. England, F. Stock, J. E. T. Gebo, K. M. Frank, A. F. Lau and E. Munson, *Journal of Clinical Microbiology*,
292 2019, **57**, e01548-01518.
- 293 4. H. Shintani, *Biocontrol Science*, 2016, **21**, 63-72.
- 294 5. V. Mugoyela and K. D. Mwambete, *Ther Clin Risk Manag*, 2010, **6**, 443-448.
- 295 6. FDA, 2004.
- 296 7. E. Pharmacopoeia, 2021.
- 297 8. WHO, 2011.
- 298 9. C. Quintelas, D. P. Mesquita, J. A. Lopes, E. C. Ferreira and C. Sousa, *Int J Pharm*, 2015, **492**, 199-206.
- 299 10. S. N. Zulkifli, H. A. Rahim and W. J. Lau, *Sens Actuators B Chem*, 2018, **255**, 2657-2689.
- 300 11. R. S. Senger, V. Kavuru, M. Sullivan, A. Gouldin, S. Lundgren, K. Merrifield, C. Steen, E. Baker, T. Vu, B. Agnor,
301 G. Martinez, H. Coogan, W. Carswell, L. Karageorge, D. Dev, P. Du, A. Sklar, G. Orlando, J. Pirkle, Jr. and J. L.
302 Robertson, *PLoS One*, 2019, **14**, e0222115.
- 303 12. J. van de Vossenbergh, H. Tervahauta, K. Maquelin, C. H. W. Blokker-Koopmans, M. Uytewaal-Aarts, D. van
304 der Kooij, A. P. van Wezel and B. van der Gaag, *Analytical Methods*, 2013, **5**.
- 305 13. P. Rostron and D. Gerber, *International Journal of Engineering and Technical Research*, 2016, **6**, 50-64.
- 306 14. Z. Li, M. J. Deen, S. Kumar and P. R. Selvaganapathy, *Sensors (Basel)*, 2014, **14**, 17275-17303.
- 307 15. K. Buckley and A. G. Ryder, *Appl Spectrosc*, 2017, **71**, 1085-1116.
- 308 16. S. Stöckel, J. Kirchhoff, U. Neugebauer, P. Rösch and J. Popp, *Journal of Raman Spectroscopy*, 2016, **47**, 89-
309 109.
- 310 17. L. M. Kandpal, B.-K. Cho, J. Tewari and N. Gopinathan, *Sensors and Actuators B: Chemical*, 2018, **260**, 213-
311 222.
- 312 18. M. K. Maruthamuthu, A. H. Raffiee, D. M. De Oliveira, A. M. Ardekani and M. S. Verma, *Microbiologyopen*,
313 2020, **9**, e1122.
- 314 19. T. N. Zu, A. I. Athamneh, R. S. Wallace, E. Collakova and R. S. Senger, *J Bacteriol*, 2014, **196**, 3983-3991.
- 315 20. D. S. Liao, J. Raveendran, S. Golchi and A. Docoslis, *Sensing and Bio-Sensing Research*, 2015, **6**, 59-66.
- 316 21. A. Sengupta, M. Mujacic and E. J. Davis, *Analytical and Bioanalytical Chemistry*, 2006, **386**, 1379-1386.
- 317 22. S. A. Strola, J. C. Baritoux, E. Schultz, A. C. Simon, C. Allier, I. Espagnon, D. Jary and J. M. Dinten, *J Biomed*
318 *Opt*, 2014, **19**, 111610.
- 319 23. P. H. Eilers, *Anal Chem*, 2003, **75**, 3631-3636.
- 320 24. M. Barker and W. Rayens, *Journal of Chemometrics*, 2003, **17**, 166-173.
- 321 25. S. Guo, P. Rösch, J. Popp and T. Bocklitz, *Journal of Chemometrics*, 2020, **34**.
- 322 26. A. Tripathi, R. E. Jabbour, P. J. Treado, J. H. Neiss, M. P. Nelson, J. L. Jensen and A. P. Snyder, *Applied*
323 *Spectroscopy*, 2008, **62**, 1-9.
- 324 27. M. Dračinský, L. Benda and P. Bouř, *Chemical Physics Letters*, 2011, **512**, 54-59.
- 325 28. M. M. Ilczyszyn, M. Ilczyszyn and M. Selent, *Journal of Molecular Structure*, 2020, **1212**.
- 326 29. N. A. Marley, C. K. Mann and T. J. Vickers, *Applied Spectroscopy*, 2016, **38**, 540-543.
- 327 30. K. S. Kalasinsky, T. Hadfield, A. A. Shea, V. F. Kalasinsky, M. P. Nelson, J. Neiss, A. J. Drauch, G. S. Vanni and
328 P. J. Treado, *Anal Chem*, 2007, **79**, 2658-2673.
- 329 31. O. N. Koroleva, L. A. Shabunina and V. N. Bykov, *Glass and Ceramics*, 2011, **67**, 340-342.
- 330 32. S. Yang, B. Li, M. N. Slipchenko, A. Akkus, N. G. Singer, Y. N. Yeni and O. Akkus, *J Raman Spectrosc*, 2013, **44**,
331 1089-1095.
- 332 33. M. Ritz, L. Vaculíková, J. Kupková, E. Plevová and L. Bartoňová, *Vibrational Spectroscopy*, 2016, **84**, 7-15.
- 333 34. E. Cloutis, P. Szymanski, D. Applin and D. Goltz, *Icarus*, 2016, **274**, 211-230.

- 334 35. M. L. Paret, S. K. Sharma, L. M. Green and A. M. Alvarez, *Appl Spectrosc*, 2010, **64**, 433-441.
335 36. C. S. Ho, N. Jean, C. A. Hogan, L. Blackmon, S. S. Jeffrey, M. Holodniy, N. Banaei, A. A. E. Saleh, S. Ermon and
336 J. Dionne, *Nat Commun*, 2019, **10**, 4927.
337 37. C. Wang, F. Madiyar, C. Yu and J. Li, *J Biol Eng*, 2017, **11**, 9.
338 38. N. Kuhar, S. Sil, T. Verma and S. Umapathy, *RSC Advances*, 2018, **8**, 25888-25908.
339 39. Y. Ryu, M. Hong, S. B. Kim, T. K. Lee and W. Park, *Journal of Microbiology*, 2021, **59**, 491-499.
340 40. W. H. Coleman, D. Chen, Y. Q. Li, A. E. Cowan and P. Setlow, *J Bacteriol*, 2007, **189**, 8458-8466.
341 41. W. H. Nelson, R. Dasari, M. Feld and J. F. Sperry, *Appl Spectrosc*, 2004, **58**, 1408-1412.
342 42. L. Kong, P. Zhang, P. Setlow and Y. Q. Li, *Anal Chem*, 2010, **82**, 3840-3847.
343 43. Q. Li, J. Yang, W. Fan, D. Zhou, X. Wang, L. Zhang, M. Huo and J. C. Crittenden, *Colloids and Surfaces B: Biointerfaces*, 2018, **162**, 35-42.
344 44. M. Carrera, R. O. Zandomeni, J. Fitzgibbon and J. L. Sagripanti, *J Appl Microbiol*, 2007, **102**, 303-312.
345 45. G. Teng, Q. Wang, X. Cui, K. Wei, W. Xiangli and G. Chen, *Journal of Raman Spectroscopy*, 2021, **52**, 805-
346 814.
347 46. S. Pahlow, S. Meisel, D. Cialla-May, K. Weber, P. Rösch and J. Popp, 2015.
348

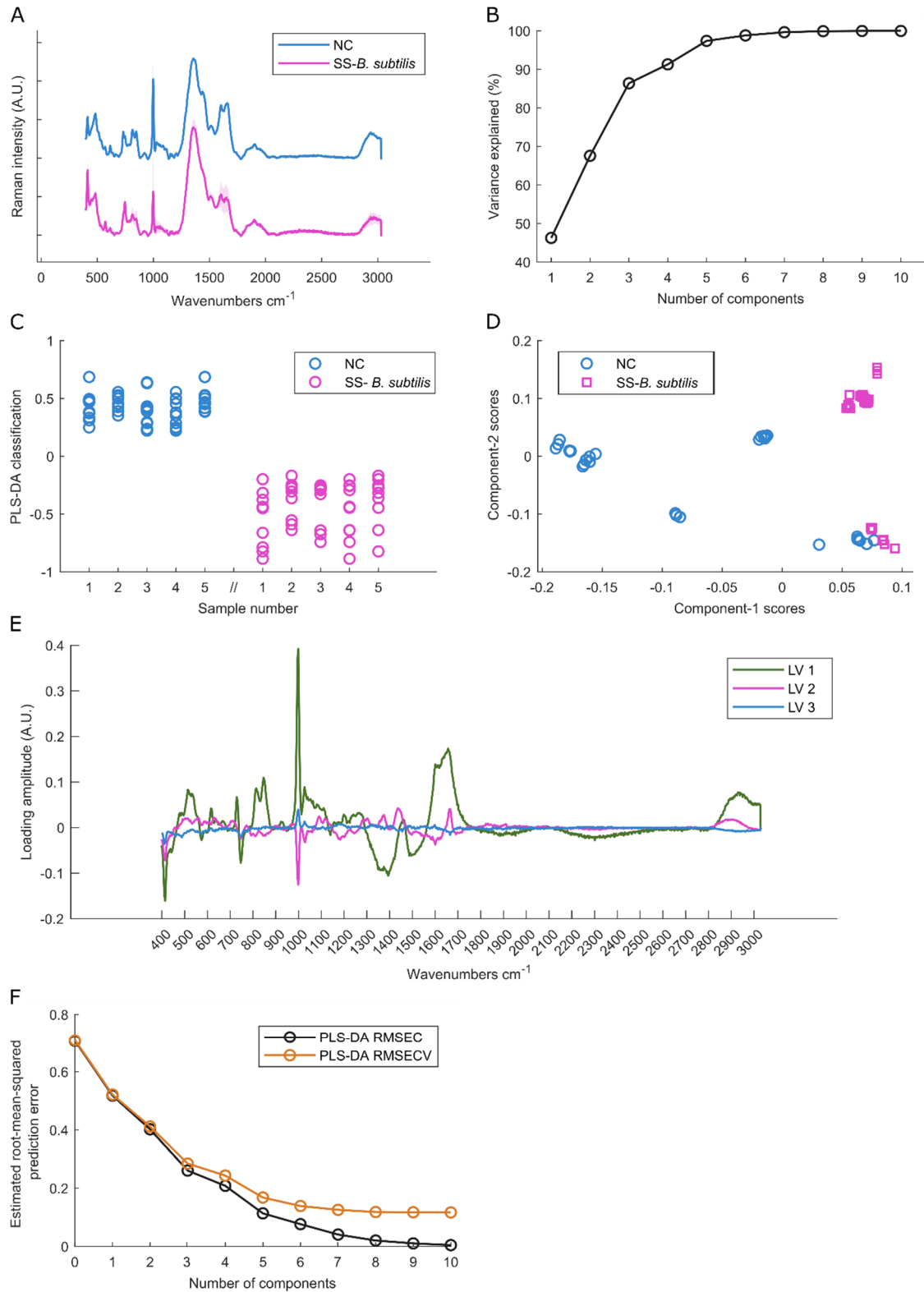


Figure 1: Detection of bacteria in drug product primary package. Raman spectroscopy in combination with partial least squared discriminant analysis (PLS-DA) to distinguish 5 spiked samples with 3×10^8 CFU/ml of vegetative *B. subtilis* (SS-vBs) from 5 negative control (NC) samples. **(A)** Baselined corrected and normalized Raman spectra from SS-vBs (magenta) and NC (blue) represented by data mean (solid line) and \pm SD (shadow). **(B)** Variance explained plot of the PLS-DA model in function of the number of components into consideration. **(C)** Classification plot after analysis with a three-component PLS-DA model. Circles represent each measurement of the spiked (magenta) and NC (blue) samples along the vial. **(D)** Score plot of the samples considering scores obtained in component 1 (X-axis) vs components 2 (Y-axis). **(E)** Loading vectors (LV) extracted from the three-component PLS-DA analysis, colors represent each LV1 (green), LV2 (magenta) and LV3 (blue). **(F)** Plot of the estimated root-mean-squared error of calibration (RMSEC, black line) and cross-validation (RMSECV, orange line) in function to the number of components added into the model.

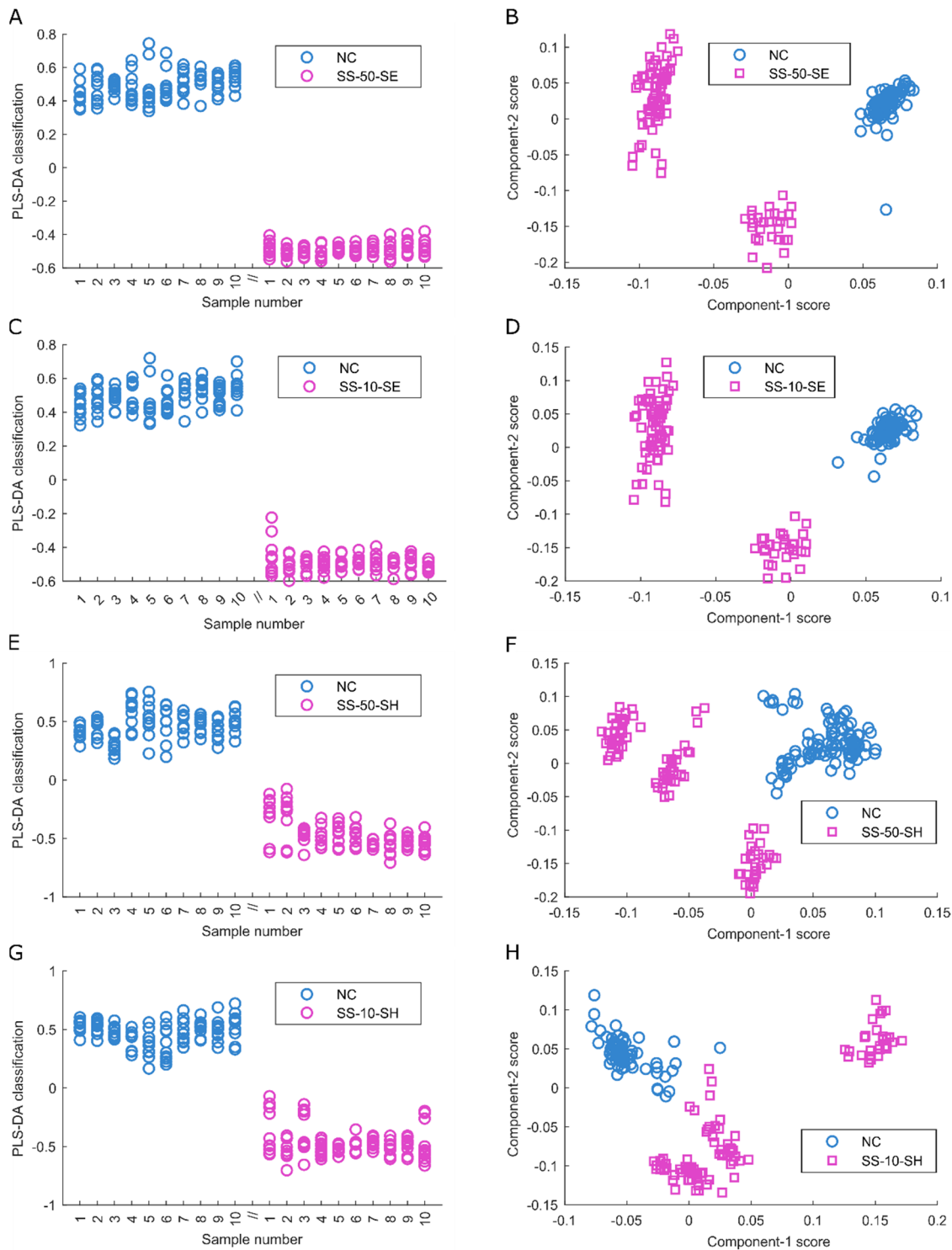


Figure 2: Detection of low bacterial concentrations in vegetative forms. Raman spectra from 10 spiked samples (SS) and 10 negative controls (NC) were collected and analyzed using a three-component partial least squared with discriminant analysis (PLS-DA) model. Classification and score plots obtained in components number 1 and 2 of **(A-B)** SS with 50 CFU/ml of *S. enterica* (SS-50-SE); **(C-D)** SS with 10 CFU/ml of *S. enterica* (SS-10-SE); **(E-F)** SS with 50 CFU/ml of *S. haemolyticus* (SS-50-SH); **(G-H)** SS with 10 CFU/ml of *S. haemolyticus* (SS-10-SH).

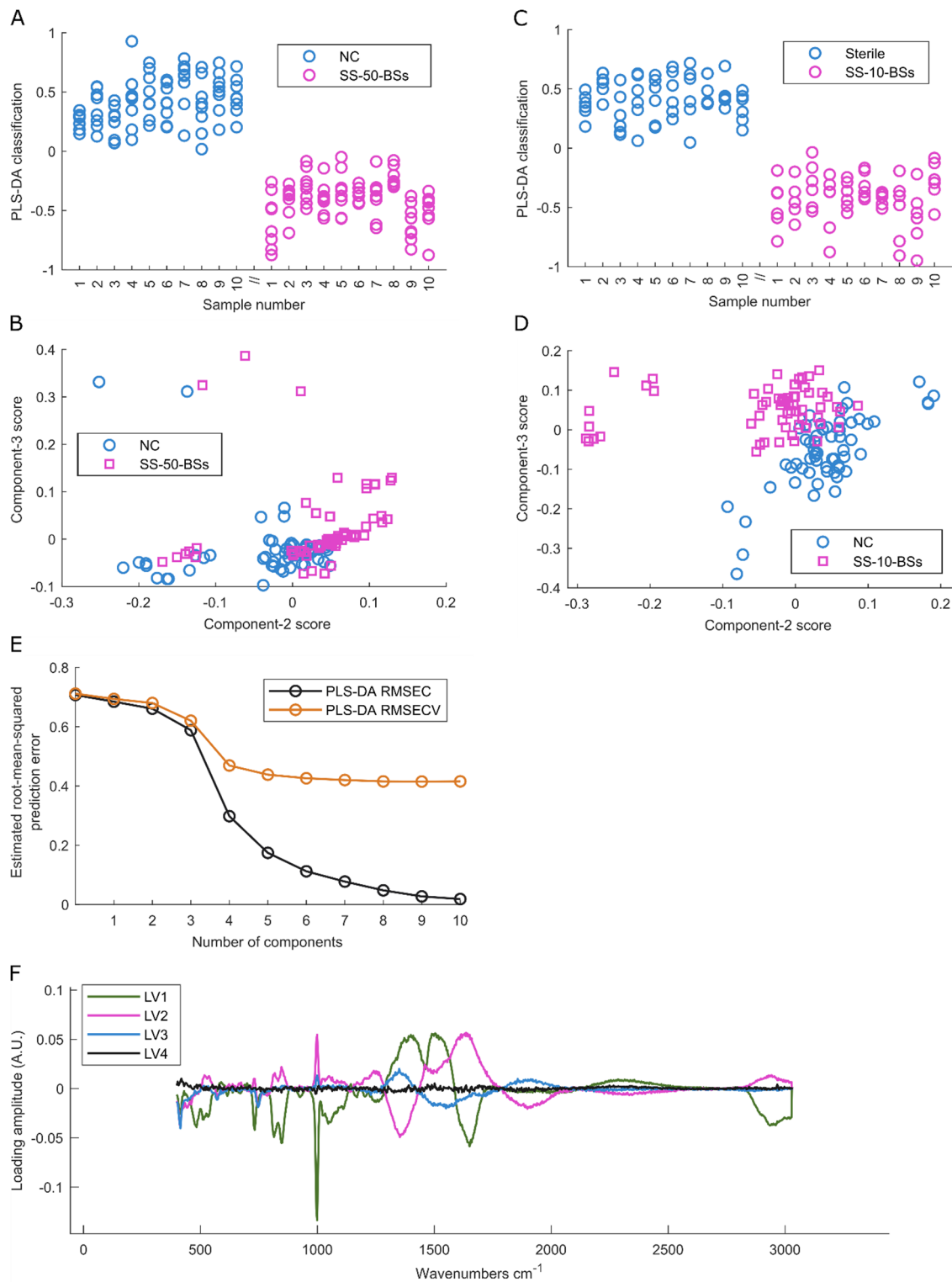


Figure 3: Detection of spores. Raman spectra from 10 spiked samples (SS, magenta) and 10 negative controls (NC, blue) were collected and analyzed using a four-component partial least squared with discriminant analysis (PLS-DA) model. Classification and score plots obtained in components number 2 and 3 of (A-B) SS with 50 CFU/ml of *B. subtilis* spores (SS-50-BSs); (C-D) SS with 10 CFU/ml of *B. subtilis* spores (SS-10-BSs). (E) Plot of the estimated root-mean-squared error of calibration (RMSEC, black line) and cross-validation (RMSECV, orange line) in function to the number of components added into the model. (F) Loading vectors (LV) extracted from the four-component PLS-DA analysis to visualize peaks related to spore discrimination, colors represent each LV1 (green), LV2 (magenta), LV3 (blue) and LV4 (black).

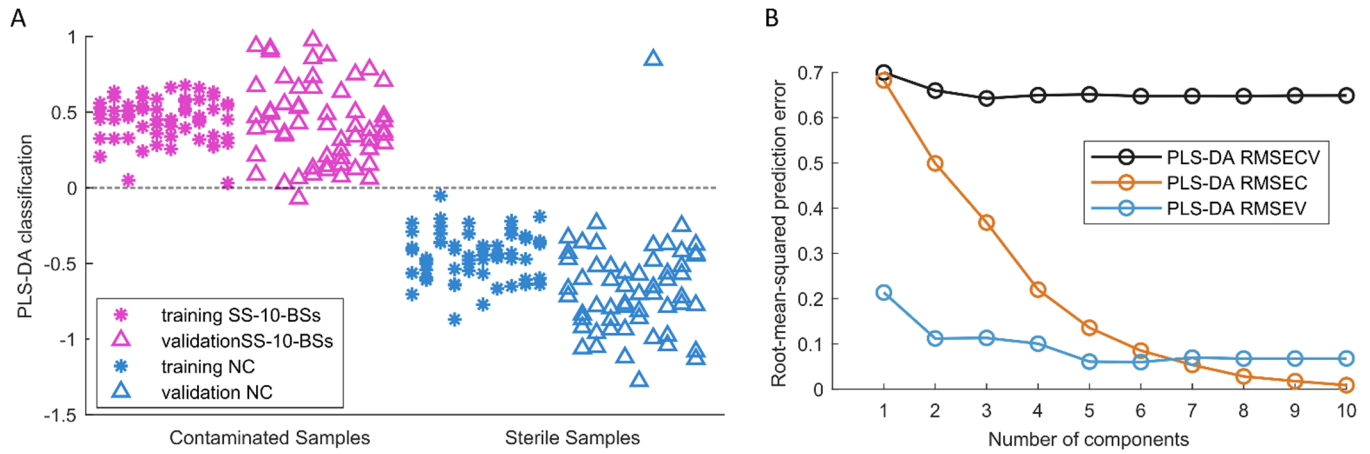


Figure 4: Model validation. Independent validation using partial least squared discriminant analysis (PLS-DA) and 10-fold cross-validation. **(A)** Classification of 10 spiked samples with 10 CFU/ml of *B. subtilis* spores (SS-10-BSs) and 10 negative controls (NC) were used as training dataset (magenta and blue stars respectively). 10 new SS-10-BSs and 10 NC samples were used as independent test set (magenta and blue triangles respectively). **(B)** Plot of the root-mean-squared error of calibration (RMSEC, orange), cross-validation (RMSECV, black line) and validation (RMSEV, blue line) in function to the number of components added into the model.