Journal of Controlled Release 367 (2024) 676–686

Contents lists available at ScienceDirect

Journal of Controlled Release

journal homepage: www.elsevier.com/locate/jconrel

Hyper-spectra imaging analysis of PLGA microspheres *via* machine learning enhanced Raman spectroscopy

Minghe Li [a,1], Ruifeng Wang [a,b,1], Quanying Bao [a,2,*]

[a] *Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT 06877, USA*
[b] *Department of Pharmaceutical Sciences, University of Connecticut, Storrs, CT 06269, USA*

ARTICLE INFO

Keywords:
Long-acting injectables
PLGA
Microspheres
Raman microscopy
Machine learning
Particle statistics

ABSTRACT

Long-acting injectables (LAI) offer a cost-effective and patient-centric approach by reducing pill burden and improving compliance, leading to better treatment outcomes. Among various types of long-acting injectables, poly (lactic-*co*-glycolic acid) (PLGA) microspheres have been extensively investigated and reported in the literature. However, microsphere formulation development is still challenging due to the complexity of PLGA polymer, formulation screening, and processing, as well as time-consuming and cumbersome physicochemical characterization. A further challenge is the limited availability of drug substances in early formulation development. Therefore, there is a need to develop novel and advanced tools that can accelerate the early formulation development. In this manuscript, a novel comprehensive physicochemical characterization approach was developed by integrating Raman microscopy and the machine learning process. The physicochemical properties such as drug loading, particle size and size distribution, content uniformity/heterogeneity, and drug polymorphism of the microspheres can be obtained in a single run, without requiring separate methods for each attribute (*e.g.*, liquid chromatography, particle size analyzer, thermal analysis, X-ray powder diffraction). This approach is non-destructive and can significantly reduce material consumption, sample preparation, labor work, and analysis time/cost, which will greatly facilitate the formulation development of PLGA microsphere products. In addition, the approach will potentially be beneficial in enabling automated high throughput screening of microsphere formulations.

1. Introduction

Long-acting injectables (LAI) (*e.g.*, aqueous drug suspension, microspheres, *in situ* forming implants) have drawn tremendous attention in recent years for drug product development. Among different types of long-acting injectables, poly (lactic-*co*-glycolic acid) (PLGA) microsphere has been the most extensively investigated and reported to date. Efforts have been made to understand the formulation and processing, develop advanced analytical characterization method, identify critical quality attributes, develop dissolution/*in vitro* release testing method as well as establish *in vitro-in vivo* correlations [1–5]. With the advancement in these areas, three generic microsphere products have been approved (2023) by the US FDA. Only one brand Risperidone microsphere (Rykindo) was approved by the US FDA recently (2023) in the past 5 years. Although the development of long-acting injectable products is slow in general due their complex nature, aqueous suspensions have gained greater momentum with more brand product approvals by the Agency, to name a few, Abilify Asimtufii (an aripiprazole suspension, Apr. 2023), Apretude (a cabotegravir suspension, Dec. 2021), Cabenuva kit (a cabotegravir and a rilpivirine suspension, Jan. 2021), Invega Hafyera (a paliperidone palmitate suspension, Aug. 2021) [6]. The momentum shift may be attributed to many aspects. Long-acting injectable aqueous suspensions require fewer steps of formulation (readily to be suspended in diluent) and processing given the drug substance meets all the required target profiles. On the contrary, microsphere products are required to be formulated into particles with spherical morphology before being suspended in diluent/media. In addition, PLGA microsphere product development is challenging in early formulation screening and optimization due to the complexity of PLGA polymer selection and laborious solid-state characterization (*e.g.*,

* Corresponding author at: Material and Analytical Sciences Department, Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, CT 06877, USA.
E-mail address: quanying.bao@alexion.com (Q. Bao).
[1] Minghe Li and Ruifeng Wang contributed equally to this manuscript.
[2] Current Address: Alexion AstraZeneca Rare Disease, 100 College Street, New Haven, CT 06510, USA.

https://doi.org/10.1016/j.jconrel.2024.01.071
Received 28 November 2023; Received in revised form 30 January 2024; Accepted 31 January 2024
Available online 9 February 2024

drug loading, particle size and distribution, content uniformity) of the microspheres. Furthermore, the amount of drug substance used in early formulation development is sparing which greatly limits the number of experiments and/or formulations in the screening process. To accelerate early formulation development of PLGA microspheres, there is an unmet need to simplify the procedures of formulation and/or characterization.

Raman spectroscopy is a technique that quantifies the inelastic scattering of photons, a phenomenon referred to as Raman scattering. Nowadays, Raman spectroscopy is not only limited to spontaneous emission and targeting vibrational modes of molecules but is also available in coherent scattering and low-frequency modes (Terahertz Raman) [7–9]. Raman spectra can be acquired non-invasively with minimum sample preparation. Samples can be stored and maintained at the native condition which can be directly used for other analytical measurements afterward. In addition to the straightforward sample preparation and instrumentation operation, Raman measurement and analysis are also exceptionally fast and reproducible, rendering it a robust and time-efficient method [10,11]. Owing to all the advantages mentioned above, Raman spectroscopy has been applied to various applications in pharmaceutical analysis. For example, Raman spectroscopy has been used to track reaction progress in real-time, discriminate polymorphs of the crystal structure, and measure the quantitative chemical composition [12–16]. Although Raman spectroscopy has been extensively applied in pharmaceutical analysis, it is not sufficient to elucidate subtle variations due to the complexity of pharmaceutical formulations [17]. To address that, scientists combine the advantages of the imaging technique with Raman spectroscopy, as a microscope system provides Raman the spatial X and Y resolution to reveal the morphology and inhomogeneity of the sample [18]. Furthermore, the confocal Raman microscopy can selectively probe the XYZ location with submicron accuracy [19,20]. Due to the broad signal range and abundant spatial information, each Raman spectrum image stores millions of spectra ranging over thousands of wavenumbers. The enormous dataset of the Raman images is inefficient and inaccurate to be analyzed by conventional dimensionality reduction methods. Especially in early-stage drug formulation development, the high-throughput screening steps require minimum sample preparation and *in situ* measurement. The substrate can significantly interfere with the samples, sometimes even being the dominant signal, not to mention the perturbation from the starting material and impurities. There is an urgent need for an intelligent method that can eliminate all the interference and perform data mining for key chemical compound signals. Thus, machine learning based data analysis method has become popular in spectrum studies [21–23].

As an emerging tool, machine learning has accelerated the research and development in pharmaceutical industry from early discovery to clinical trials [24,25]. Machine learning bridges the gap between innovative technology and drug development. With the development of early-stage drug discovery, clinical trial research, quality control methods, and smart portable devices, pharmaceutical study tends to transform into data-driven research [24–26]. Machine learning is one of the best solutions to manage big data. When it comes to the Raman hyperspectral images, the variation in spectrum intensity and spectral profiles, due to the sophisticated microenvironment of the molecules within the formulation, complicates the conventional linear regression-based methods. However, the machine learning-based algorithm approach is able to better capture, recognize and classify the spectra even with microenvironment perturbation or missing peaks from the raw data.

Raman imaging technique is an ideal method for the physicochemical characterization of microspheres in that, *1)* the particle size range of microsphere formulations (in the range of 1–100 μm for most the microsphere products) falls within the sweet spot of the resolution of the optical microscope (~1 μm); *2)* the spherical morphology of microspheres facilitate their identification under the microscope; and *3)* the translucent polymer (PLGA) allows light to transmit and reach the drug molecules underneath the polymeric shell. Here, a machine learning algorithm was introduced to the label-free Raman spectral imaging, enabling chemical discrimination. The whole process (Fig. 1 a) comprises three major steps: Raman spectrum collection, machine learning model building, and particle statistics. The demonstrated machine learning enhanced Raman imaging can achieve comparable results with multiple conventional analytical methods.

Two molecules (Risperidone and medroxyprogesterone acetate (MPA)) with distinct solubility (over 100 μg/mL of Risperidone *vs.* less than 5 μg/mL of MPA) [27,28] and the needs for LAI formulations have been selected as models to perform the studies. Risperidone is an antipsychotic drug used to treat schizophrenia, bipolar disorder, and irritability associated with autism. Risperidone microspheres (the first brand product Risperdal Consta) have been extensively investigated to understand the polymer, formulation attributes as well as *in vitro-in vivo* correlations and hence facilitate the generic approval in the past several years [29–32]. With these efforts, one generic (one of the first three generic microspheres) and one brand risperidone microspheres have been approved in 2023. MPA is a hormone progestin for contraception use. As compared to Risperidone, MPA has a much lower solubility and therefore its commercially available products are LAI drug suspensions (*e.g.*, Depo Provera CI, Depo-subQ Provera 104). However, it is unknown if MPA can be formulated into LAI microsphere formulations, representing a plethora of new compounds in early phase formulation screening. It is meaningful to understand if machine learning enhanced Raman imaging approach can be applied to the two small model molecules. The results revealed the possibility of using the Raman microscope to perform small volume, noninvasive, versatile physicochemical characterization for PLGA microspheres in early-stage formulation development.

2. Materials and methods

2.1. Materials

The commercially available compounds, Risperidone, medroxyprogesterone acetate (MPA), PLGA (Resomer® RG 752H, acid terminated, lactide: glycolide 75:25, Mw 4000-15,000), poly (vinyl alcohol) (PVA, with an average molecular weight of 30–70 kDa), and trifluoroacetic acid (TFA) were obtained from Sigma-Aldrich (St. Louis, MO, USA). Dichloromethane (DCM), acetonitrile (HPLC grade), and water (HPLC grade) were purchased from Fisher Scientific (Pittsburgh, PA, USA). Polyvinylpyrrolidone (PVP, Kollidon® VA64) was obtained from BASF (Mount Olive, NJ, USA). Milli-Q® water was used for all studies. All other reagents used in this research were of analytical grade.

2.2. Methods

2.2.1. PLGA microsphere preparation

MPA and Risperidone microsphere formulations with different drugs were prepared using an emulsion solvent evaporation method (Fig. 1d). Briefly, the PLGA was dissolved in DCM (14.3%, *w*/w). A predetermined amount of drug powder was added to the PLGA solution when the polymer was fully dissolved. PVA solution (5 mL, 1%, *w/v*, saturated with DCM) was then added into the PLGA/drug solution and homogenized at 3600 rpm for 2 min (IKA Work, Inc., NC, USA) to form the primary oil-in-water emulsions. The resulting emulsions were transferred into 0.1% *w/v* PVA at room temperature under a fume hood and stirred at 450 rpm for 2 h. Then further organic solvent removal was conducted under a vacuum (*ca.* 150 mbar) at room temperature for 2 h. The solidified microspheres were then washed with water three times, passed through a 212 μm sieve in dispersed form, and subsequently freeze-dried.

2.2.2. Amorphous solid dispersion preparation for MPA

Amorphous solid dispersions were used to study the ability of Raman imaging methods to quantify the amounts/ratios of different

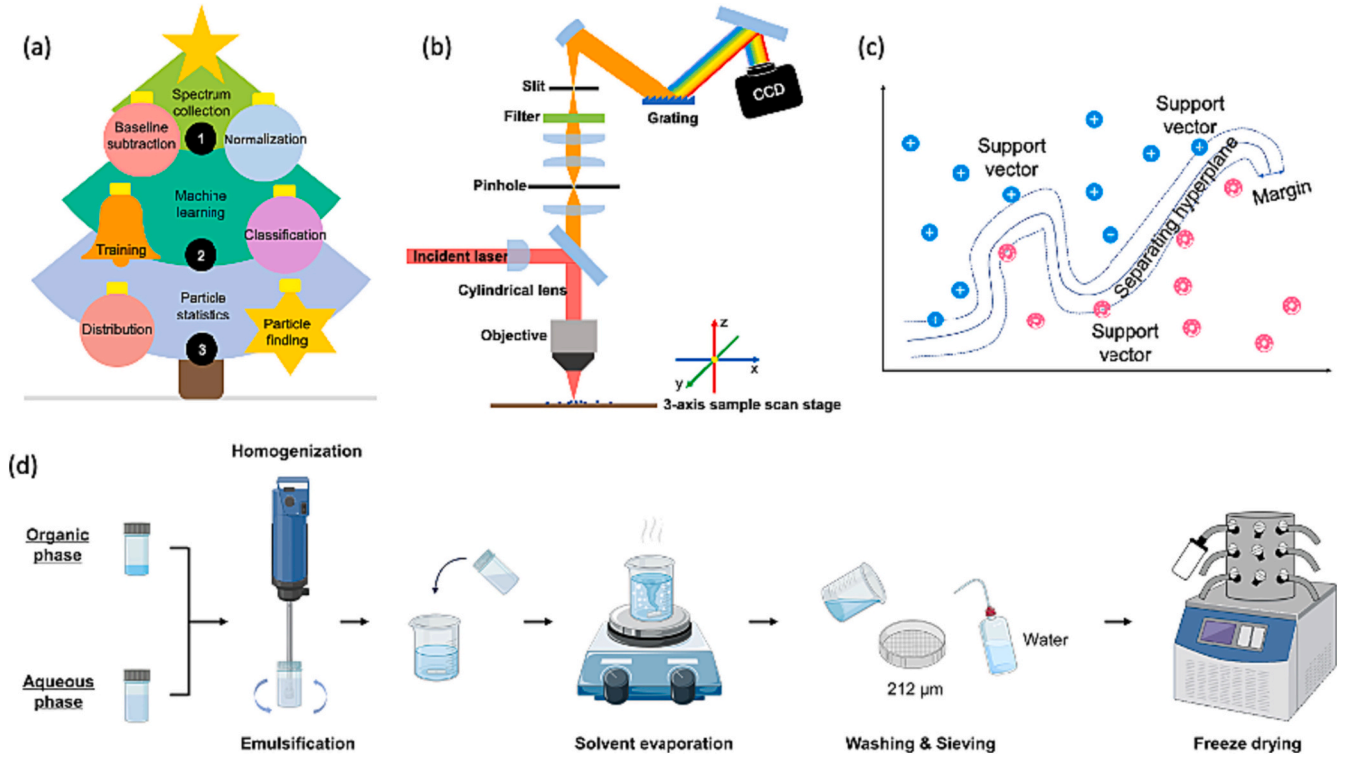

Fig. 1. (a) Workflow of the machine learning enhanced data acquisition and processing. (b) Line focus Raman imaging system. The cylindrical lens elongates the point focus into line focus to improve the throughput of the spectrum collection. (c) Nonlinear kernel support vector machine (SVM). (d) Schematic drawing of microsphere preparation process.

polymorphs (amorphous or different crystalline forms) within microsphere samples. Briefly, the MPA (10 mg) was dissolved in 1 mL DCM to make a drug solution and PVP (40 mg) was dissolved in 1 mL DCM to make a polymer solution. The two solutions above were mixed and stirred at 300 rpm and 50 °C to evaporate DCM. The solid was then vacuum-dried at room temperature overnight.

2.2.3. Ultra-high performance liquid chromatography (UPLC) analysis

The concentration of MPA and Risperidone was determined using a UPLC system (Acquity, Waters, USA). An Acuqity® Premier HSS T3 column (2.1 mm × 50 mm, 1.8 μm, Waters®) was used as the stationary phase and the column temperature was set at 30 °C. Different mobile phases and elution methods for each drug are as follows. For MPA, the mobile phase (a mixture of acetonitrile and water (65:35, *v*/v)) was used in an isocratic condition at a flow rate of 0.5 mL/min for 2.5 min. The injection *v*olume was 1 μL and the detection wavelength was set at 244 nm. For Risperidone, the mobile phase (a mixture of 0.1% *v*/v trifluoracetic acid in water and acetonitrile (95:5 *v*/v) (Mobile phase A) and 0.1% trifluoracetic acid in water and acetonitrile (5:95, v/v) (Mobile phase B) was used in a gradient condition (95% A to 5% A in 2.5 min) at a flow rate of 0.5 mL/min for 4 min. The injection volume was 1 μL and the detection wavelength was set at 275 nm. The chromatograms were then analyzed using Empower® 3.0 software.

2.2.4. Drug loading determination using UPLC

The prepared microspheres (~5 mg) were weighed and dissol*v*ed in 10 mL acetonitrile to determine the drug loading. The samples were sonicated in a water bath for 15 min. Then the samples were diluted with a solution of acetonitrile and water (65:35, *v*/v), followed by UPLC analysis as described above. All experiments were performed in triplicate and the results were expressed as the mean ± SD. The drug loading was calculated as:

Drug loading (%) = Weight of drug-loaded / Weight of microspheres * 100%.

2.2.5. Particle size

Particle size and size distribution of the prepared microspheres were determined using an AccuSizer A7000APS optical particle size analyzer (Engeris Inc., Port Rickey, FL, USA). The prepared microspheres were suspended in 0.1% *w/v* PVA solution at a concentration of 2 mg/mL. Approximately 100 μL of samples were injected into the particle size analyzer for data analysis. The data was collected using AccuSizer particle sizing analysis software (Engeris Inc., Port Rickey, FL, USA).

2.2.6. Polarized light microscope (PLM)

The microsphere samples were visualized using an Olympus BX51 microscope (Olympus Optical, Tokyo, Japan) equipped with a polarized light filter and a Zeiss Axiocam 305 colour camera. An aliquot of the samples was loaded on a glass slide followed by applying a drop of mineral oil on the samples. The samples were then covered by a cover glass for visualization. The PLM images were captured with a 10-fold magnification objective lens. All the images were collected and analyzed using ZEN software (version 3.3 blue edition, Zeiss, Germany).

2.2.7. Differential scanning calorimetry (DSC)

Thermal analysis of the samples was conducted using Discover differential scanning calorimetry (DSC 2500, TA instruments, New Castle, DE, USA). Approximately 1 mg of samples were weighed into standard hermetic aluminum pans and sealed with an aluminum lid. The samples were equilibrated at 30 °C and then heated to 300 °C at a ramp rate of 20 °C/min. Nitrogen was used as purging gas at a flow rate of 50 mL/min. The data collection and analysis were performed using TRIOS software (TA instruments, New Castle, DE, USA).

2.2.8. X-ray powder diffraction (XRPD)

XRPD patterns were collected using a Bruker D8 ADVANCE DAVINCI diffractometer equipped with LYNXEYE detector (Bruker, Billerica, MA, USA). The X-ray was generated by sealed Cu tube for Cu Kα (1.54178 Å) at 40 kV and 40 mA. XRPD collection was performed through 2 theta

range between 2° to 35° with step of 0.05°/*sec* and 0.5 s per step. The XRPD data was analyzed by DIFFRAC EVA software (Bruker, Billerica, MA, USA).

2.2.9. Raman microscopy

Raman spectra were acquired using a Renishaw inVia™ confocal Raman microscope (Renishaw, Hoffman Estates, IL, USA). The schematic is shown in Fig. 1b. The system is equipped with both 785 nm and 532 nm continuous laser sources. The laser beam is focused by a 50× long working distance objective (Leica N PLAN L 50× NA 0.50). The Raman signal is collected from epi-direction through the same objective. The backscattered signal is dispersed using a 1200 L/mm grating onto a Charged Coupled Device (CCD) resulting in a 1 cm^{-1} spectrum resolution. The Raman spectrum images were collected using a special line focus mode with 785 nm laser. The inVia™ system provides an option to intercalate one cylindrical lens into the beam path, therefore forming a line shape focus orthogonal with the scan direction on focal plane. The line focus mode significantly reduces the probability of photo damage by extending the focal volume and improving the throughput when fully utilizing the laser power. The focus of the beam was set to live track the top surface of the microsphere to avoid signal loss of the large particles. Each spectrum image covers 500 × 500 pixels with the scan step size of 2.8 μm (field-of-view area 1400 μm × 1400 μm). After image acquisition, comic ray removal and baseline subtraction operations were performed by the built-in function of the inVia™ system. After epi-direction collection and optical grating, the real conjugate plane of the focal plane is projected onto the CCD array with one axis encoded with spectrum information and the second axis encoded with spatial information.

2.2.10. Machine learning model

The machine learning was performed by Classification learner app implemented in MATLAB R2021b (MathWorks, MA, USA). The spectrum range used for training is from 603 cm^{-1} to 1725 cm^{-1}. Six predictive models were used: the decision trees, the support vector machine (SVM), the nearest neighbor (KNN) classifiers, ensemble classifiers, the neural network classifiers, and the naïve Bayes classifiers. Among all the models, the accuracy, training, and prediction time of the SVM models are generally better than other categories. SVM maps training examples to points in space to maximize the width of the gap between the categories. As shown in Fig. 1c, the nonlinear kernel SVM algorithm could very preciously distinguish the two categories.

3. Results and discussion

3.1. Model construction

Training set spectra were collected using the pure compound of the PLGA, APIs (*e.g.*, MPA, Risperidone), and glass slide. Each training data set contains 50 × 50 pixels of Raman spectrum with 50 mW laser power and 1 s line focus exposure time. Each training set including the corresponding API, PLGA, and glass was generated for two individual microsphere formulations. The misclassification cost is defined by the ratio of the API and glass Raman cross-section which can be estimated by pure compound reference spectrum. The user-defined misclassification cost aims to minimize the strong substrate interference. The model was then trained and selected based on performance. Models were built successfully with good accuracy using both MPA and risperidone. The modeling results of MPA are shown in Fig. 2 as a representative. As

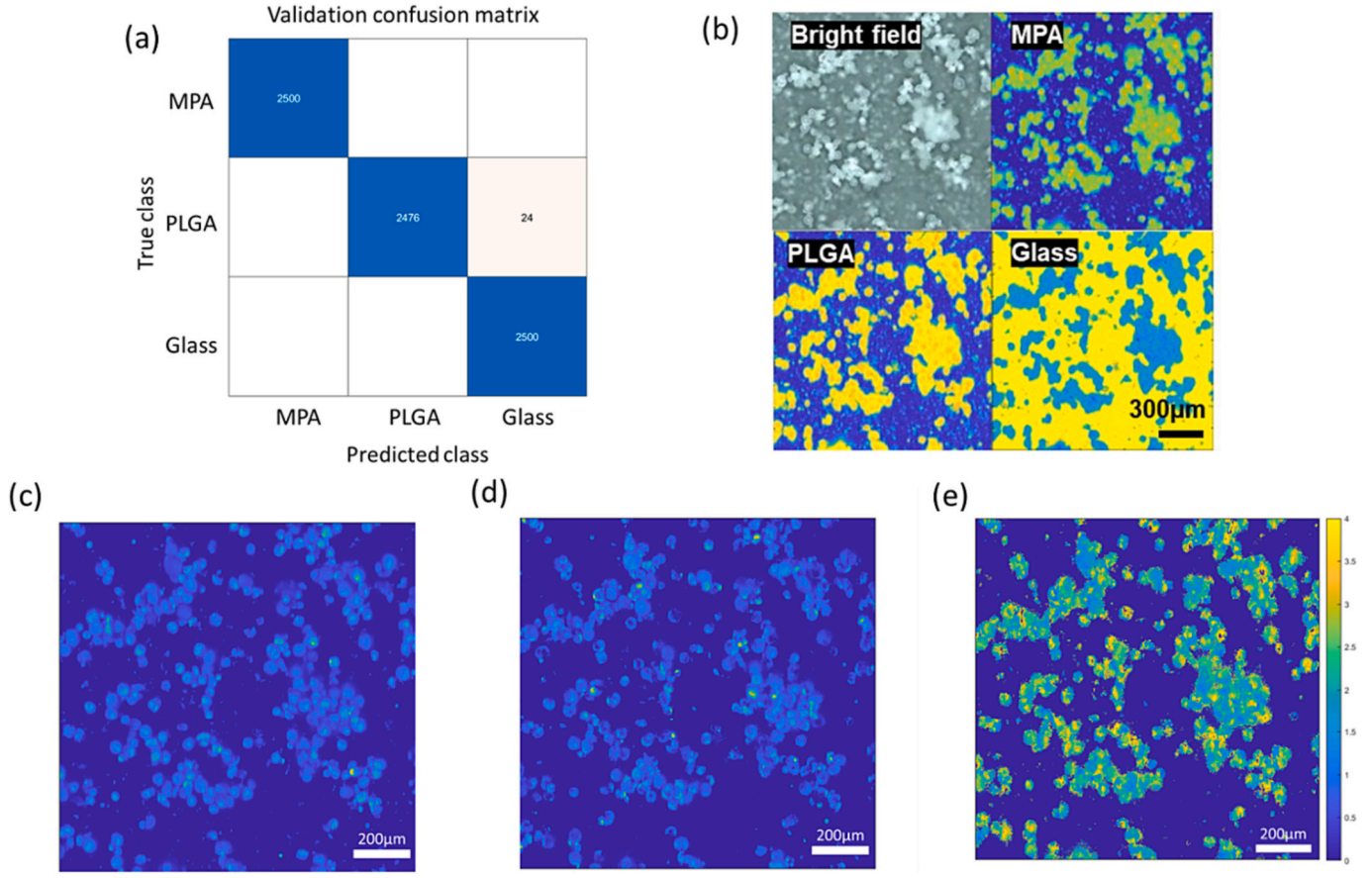

Fig. 2. (a) Confusion matrix of the Coarse Gaussian SVM model. The overall accuracy is larger than 99% for the three components. (b) Brightfield image and machine learning classified spectrum images of the representative MPA microspheres. Quantitative spectrum intensity map of the MPA (c) and PLGA (d) in the field-of-view. (e) The ratio of the MPA intensity over PLGA intensity.

shown in the validation confusion matrix, the overall accuracy is 99.7% in Fig. 2a. Though the overall accuracy is high, the practical difficulty of applying the model to a real dataset lies in the significant covariance between APIs with PLGA. In most cases, the APIs and PLGA are colocalized which causes confusion for the algorithm to assign the class. In this situation, the analysis for the microsphere is no longer a binary classification. To best predict microsphere samples, the data processing relies on the machine learning classification score shown in Fig. 2b.

Once the optimal model is chosen, the algorithm assigns triplet classification map to each chemical component of the formulation. If one pixel is assigned as glass, it is considered as background and excluded from the total field-of-view drug loading calculation. Other pixels can be assigned as APIs, PLGA or both indicating free drug, empty microsphere, or encapsulated microsphere. The quantitative analysis fuses the classification result with the spectrum intensity of the pixel. The spectrum intensity maps of the APIs and PLGA were produced by matrix multiplication, using the raw data matrix and the normalized zero-centered reference spectrum of the APIs and PLGA.

The spectrum intensity map reveals the spatial distribution of both the MPA and PLGA within the microsphere structure. Fig. 2c and Fig. 2d vividly illustrates the clustering and heterogeneity of the microsphere, a result of the precise auto-focus tracking provided by the inVia™ system. Besides the free drug crystal fragment in the background, the microspheres are all in uniformly round shape. Meanwhile, inter-sphere heterogeneity of the microsphere sample can be observed through the ratio map between MPA and PLGA spectrum intensity map in Fig. 2e.

3.2. Quantitative evaluation

The proper mapping of the Raman signal to its corresponding chemical composition is the most critical for the quantitative analysis. Microsphere samples with a series of drug loadings were prepared (Table 1) for evaluation of the linearity of the Raman derived drug loading *versus* the UPLC measurement. Quantification of different materials through Raman is based on the following equation [33,34],

$$I = CNI_0\sigma\Omega lF(T)$$

where I is Raman signal intensity, C is a constant, N is the number density, I_0 is the laser intensity, σ is the Raman cross-section of the molecule, Ω is the scattering solid angle, l is the beam path, and $F(T)$ is the temperature dependent factor. For the measurement of the same instrument and experimental parameters, the variables besides the N are all constant. So that the Raman intensity I is only linearly related to the number density N. As the volume of the sample for data collection is the same, the mass of drug or polymer is proportional to the intensity of Raman signal.

The drug loading from Raman imaging is calculated through two steps:

a) the quantitative spectra of the API and PLGA (Q_{API} and Q_{PLGA}) equals to Raw microsphere spectra ($R_{microsphere}$) times zero-centered reference API and PLGA spectra (R_{API} and R_{PLGA}):

$$Q_{API} = R_{microsphere}{}^{*} R_{API}$$

$$Q_{PLGA} = R_{microsphere}{}^{*} R_{PLGA}$$

b) The drug loading from Raman imaging is the ratio of the area under the curve (AUC) of the API quantitative spectrum over the sum of the API and PLGA quantitative spectrum.

$$\text{Drug loading } (\%) = \text{AUC}(Q_{API})/(\text{AUC}(Q_{API}) + \text{AUC}(Q_{PLGA}))^{*}100\%$$

The drug loading of the prepared microspheres was compared between UPLC measured and their theoretical drug loading (Fig. 3a). In addition, the drug loading was also compared between the UPLC and the developed Raman imaging method (Fig. 3b). All the regression showed acceptable linearity between the drug loading results measured using UPLC and Raman imaging methods.

Free drug crystals could be observed for all the prepared MPA microspheres M1-M6 (Fig. 4), indicating some extent of drug leakage and precipitation during preparation. The free drug crystals were included in the drug loading determination using UPLC. Therefore, the drug loading of encapsulated MPA microspheres (excluding free drug) should be lower than the UPLC results. In the process of quantifying drug loading using Raman imaging, only encapsulated MPA microspheres were utilized, as the free drug crystals were out of focus during Raman imaging acquisition. These could account for the evident discrepancy between the UPLC and Raman measurements of drug loading in MPA microspheres. For the Risperidone microspheres, the drug loading determined using UPLC demonstrated good linearity with the theoretical drug loading (Fig. 3a). However, the UPLC measurements were slightly lower than the theoretical values. This could be due to the drug loss at different preparation steps. It has been reported that Risperidone might act as a catalyst for PLGA degradation during microsphere preparation [32]. In addition, further drug leakage could happen during the final washing steps. Unlike the MPA microspheres, fewer disagreements were observed between the UPLC and Raman drug loading measurements for the prepared Risperidone microspheres. The good linearity (R^2 = 0.9949) suggests a strong agreement between UPLC and the Raman imaging results (Fig. 3b). This could be attributed to better drug encapsulation of Risperidone microspheres as no free drug crystals were observed in the prepared formulations (except R6 with extra high drug loading) (Fig. 4). The downside of Raman imaging-based drug loading was the low sensitivity. The limit of the detection of drug loading using the Raman imaging can be calculated by the absolute value of the intercept over the slope of the curve, resulting in 19.8% for MPA and 6.6% for Risperidone, respectively. In the formulation development,

Table 1
The yield and drug loading of prepared microsphere formulations.

Drug	Formulation	Yield (%)	Theoretical Drug loading (%, *w*/w)	Drug loading (%, w/w) from UPLC*	Drug loading (%, w/w) from Raman
MPA	M1	62.98%	14.89%	15.45%	19.97%
	M2	71.32%	24.53%	23.75%	27.13%
	M3	75.48%	35.48%	35.31%	32.33%
	M4	81.11%	44.44%	42.40%	29.04%
	M5	69.55%	54.55%	51.76%	54.28%
	M6	77.86%	64.29%	62.09%	70.27%
Risperidone	R1	50.65%	14.29%	12.23%	12.85%
	R2	50.82%	25.00%	20.11%	18.98%
	R3	61.78%	34.65%	23.30%	26.51%
	R4	56.55%	44.54%	36.08%	36.18%
	R5	55.52%	54.55%	40.17%	41.14%
	R6	49.88%	64.71%	43.28%	44.31%

* Theoretical drug loading (%) = weight of raw API / total weight of raw API and PLGA *100%.

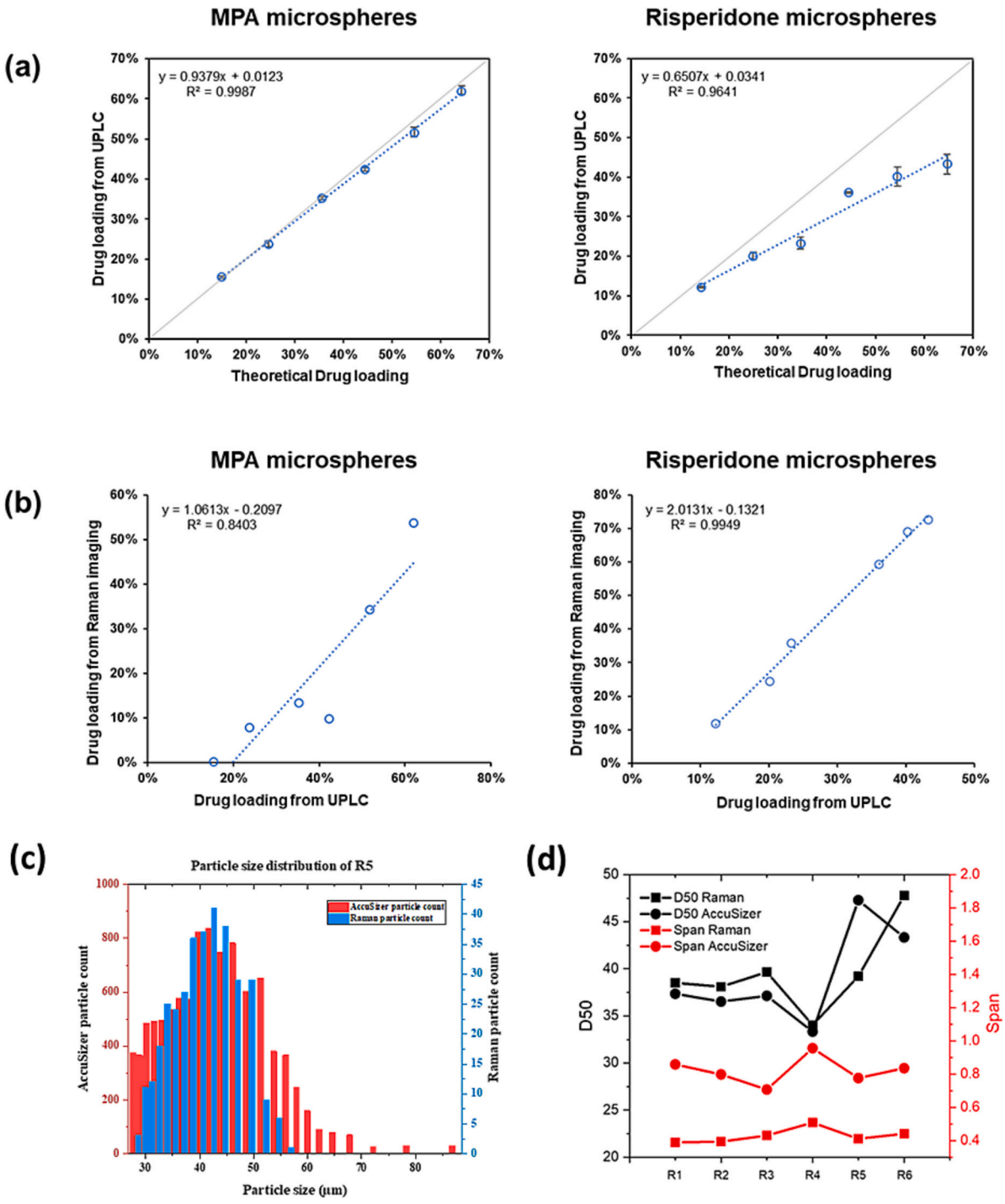

Fig. 3. (a) The theoretical drug loading *versus* experimental value from UPLC for MPA and Risperidone microsphere formulations, with the one-to-one correspondence plotted as a solid gray line. (b) The correlation between drug loadings from UPLC and Raman imaging method. (c) Particle size distribution of the formulation R5 measured by Accusizer and Raman imaging. (d) Comparison of particle size obtained using Raman imaging and AccuSizer.

calibration curve of the new formulation (using at least three standard samples for linear fitting) needs to be established as the Raman cross-section of the different API molecules and PLGAs are different. To fit for a more reliable calibration curve, at least 5 concentrations are recommended.

Other factors should be considered to quantify the drug loading through Raman imaging. First, the detection volume is a key factor of good quantification. In the current measurements, the Raman microscope was running in a confocal mode which rejects the signal out of the focal plane to maintain a constant detection volume for each pixel. Therefore, the microspheres within 10–100 μm range resulted in good signal whereas the signal was weak or out of focus for the microspheres outside of the range. Second, efforts have been made to form a single layer of microspheres during the sample preparation step to gain better results. It is noteworthy that the microscope focus tracking program will always image the top layer of the microsphere if it is not significantly far away from the pre-set z-position when encountering the overlapped microspheres. Lastly, penetration depth of the 785 nm laser in PLGA microsphere is much better than other solid samples as the laser can generate descent signal from a microsphere with a thickness of 30–40 μm.

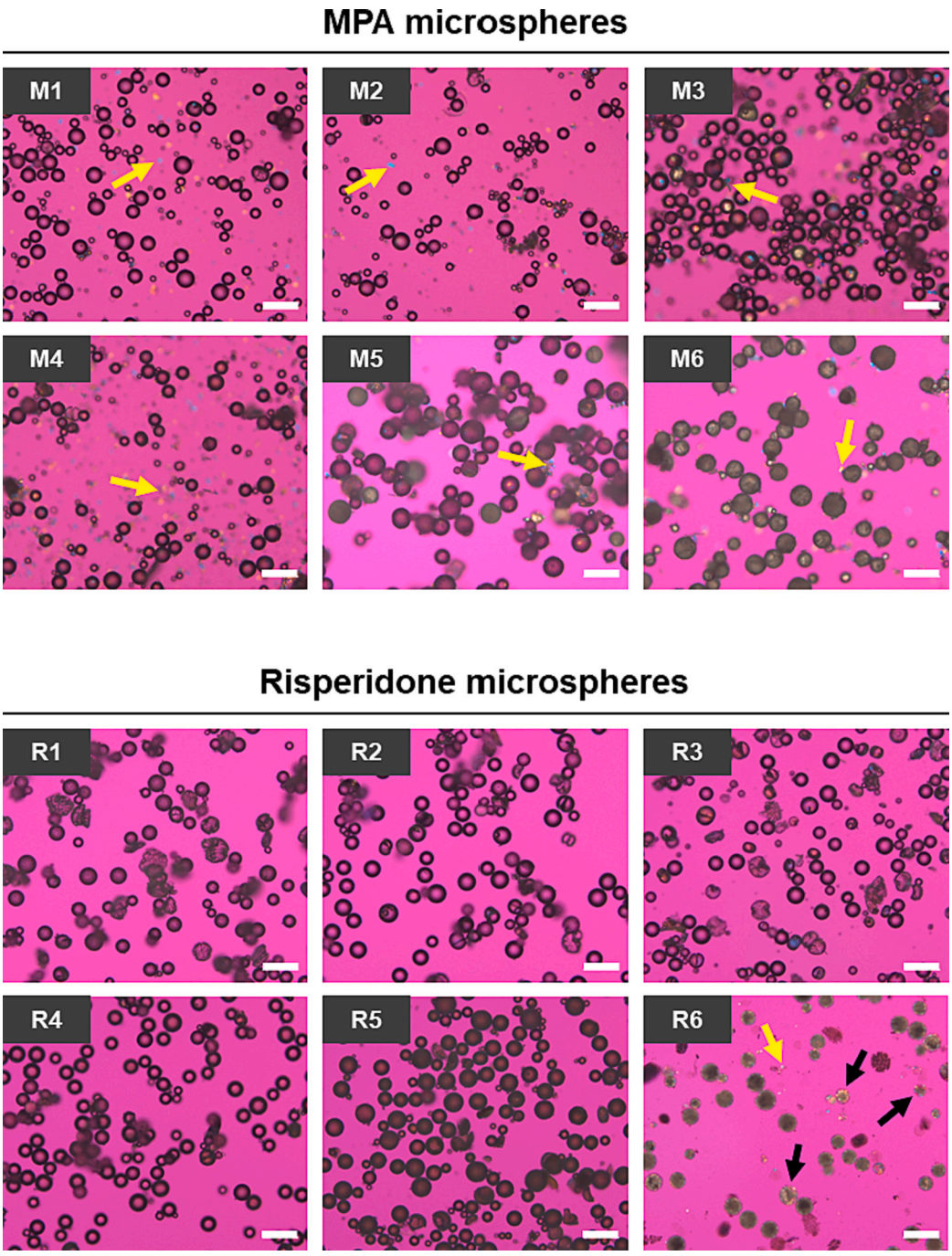

Fig. 4. PLM images of prepared microspheres. All scale bars are 100 μm. The yellow arrows point to the crystals of free drugs outside the microspheres. The black arrows point to the heterogeneous distribution of drug crystals inside microspheres. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

To understand the sensitivity of Raman imaging for particle size analysis, the particle size of the prepared microspheres was also performed by a light obscuration method (AccuSizer). It is important to note that the measurement range of the two methods is intrinsically different. The measurement range of AccuSizer is defined by the sensor of instrument. AccuSizer includes both extinction and scattering detectors to measure particles in liquid. In the current study, extinction mode was used to measure the particles with a size ranging from 1.5 to 500 μm. In comparison, Raman imaging is limited by its resolution and field-of-view. For example, the size distribution of the formulation R5 in Fig. 3c ranges from 10 to 80 μm due to the 2.8 μm resolution and 1400 μm field-of-view of Raman imaging. Although the particle counts are not identical, the relative size distributions of the two measurement methods agree with each other. Additionally, as shown in Fig. 3d, the median particle size (D50) of Risperidone microspheres (R1-R6) determined using Raman imaging closely aligned with the results from AccuSizer. However, there were noticeable variations in the span values between the two methods. This numerical discrepancy could be attributed to the wider size range of AccuSizer compared to Raman imaging as mentioned above. Furthermore, AccuSizer analyzed over ten thousand particles whereas the Raman imaging considered only several hundred particles, contributing to the disparities between the two approaches. Despite these differences, the overall trend of the mean particle size and span values remained consistent.

3.3. Covariance between particle size and drug loading of the MPA microspheres

As discussed in above Section 3.2, particle size and drug loading are two critical formulation attributes during the formulation development. In general, the drug loading and particle size were determined by liquid chromatography and particle sizer analysis, respectively. To the best of our knowledge, the particle size and drug loading have been characterized as stand-alone attributes in previous reports of PLGA-based microspheres. There has been no previous report on the characterization of drug loading within each individual microparticle, as well as the relationship between particle size and drug loading of the microspheres. The effect (covariance) between particle size and drug loading may present a better understanding of the performance of different microsphere formulations.

The drug release mechanism of PLGA microspheres has been well recognized to be a combination of polymer degradation, bulk erosion, and drug diffusion [35]. PLGA degradation (hydrolysis reaction) can be catalyzed by acids or bases. Because of the internal carboxylic acid end group, the PLGA hydrolysis can be autocatalyzed in the presence of water. It has been reported that polymer degradation was size-dependent, exhibiting more accelerated degradation in the larger particles [30]. In addition to particle size effect on the polymer degradation and/or erosion, particle size could also affect the drug release during the initial hydration process. For hydrophilic drug loading microspheres such as minocycline microspheres, smaller microspheres demonstrated a much higher release rate as the initial hydration process was faster due to the higher surface areas [5]. Lastly, drug release of PLGA microspheres was also dependent on the drug dissolution and/or diffusion rate, which directly proportional to the drug loading gradient of the microsphere. For example, two microspheres of the same size but with different drug loading could demonstrate dramatically distinctive release profiles. Therefore, it is meaningful to characterize the effect of particle size and drug loading other than their individual values.

There are many optical and imaging methods that are suitable for characterization of microparticles. For example, optical microscopy enables observation of particle morphology and size at micrometer resolution. Focus ion beam scanning electron microscopy (FIB-SEM) can visualize the microstructure (pores, drug, and PLGA polymer) of individual microspheres [2]. Light scattering technique or Coulter counter are popular for their high throughput capability however they lose spatial information when enabling statistical analysis. Raman imaging demonstrates its superiority to the other techniques mentioned above in that: *1)* it delivers moderate spatial resolution and moderate throughput with plentiful chemical information; and *2)* Raman imaging has the ability to combine size information with drug loading. As shown in Fig. 5a, the recognized microspheres from the MPA Raman intensity map have been labeled by red circle. The corresponding particle size and drug loading information are plotted in Fig. 5b. The distribution of both the size and drug loading followed the Boltzmann distribution in the selected field-of-view. In addition, it was observed that low drug loading is more possible to be found in smaller-sized microspheres. It is revealed that there was a covariance effect between drug loading and particle size in the prepared microspheres. The particle-based size and drug loading analysis will provide an intuitive understanding of the microspheres, as well as insights into the performance of the final formulations.

3.4. Amorphous content determination of the MPA microspheres

The pharmaceutical solid-state forms have been proven to greatly affect the chemical stability, dissolution profile, and formulation hardness of the final drug product. For the solid microsphere particles, it is necessary to understand the solid state of the drug particles inside the prepared microparticles. It is rare that the crystal form changes during the preparation process as the most stable form is selected for the formulation development. However, the crystal drug substance may be plasticized by the solvent, aqueous media as well as PLGA polymer, leading to amorphization of the drug during the process. Therefore, it is necessary to characterize the amorphous and crystal composition in the prepared microspheres. The analysis of the amorphous and crystal forms can be done by XRPD, DSC, and solid-state nuclear magnetic resonance (ssNMR). Optical methods, such as mid-IR, terahertz spectroscopy, and Raman, can provide fast and sensitive measurements of the composition of amorphous and crystal APIs. This is due to their sensitivity towards low energy vibrational and rotational modes.

In the present study, the amorphous content can be qualitatively and quantitatively determined by simply adding the amorphous reference spectrum of MPA (with known amorphous content) to either the

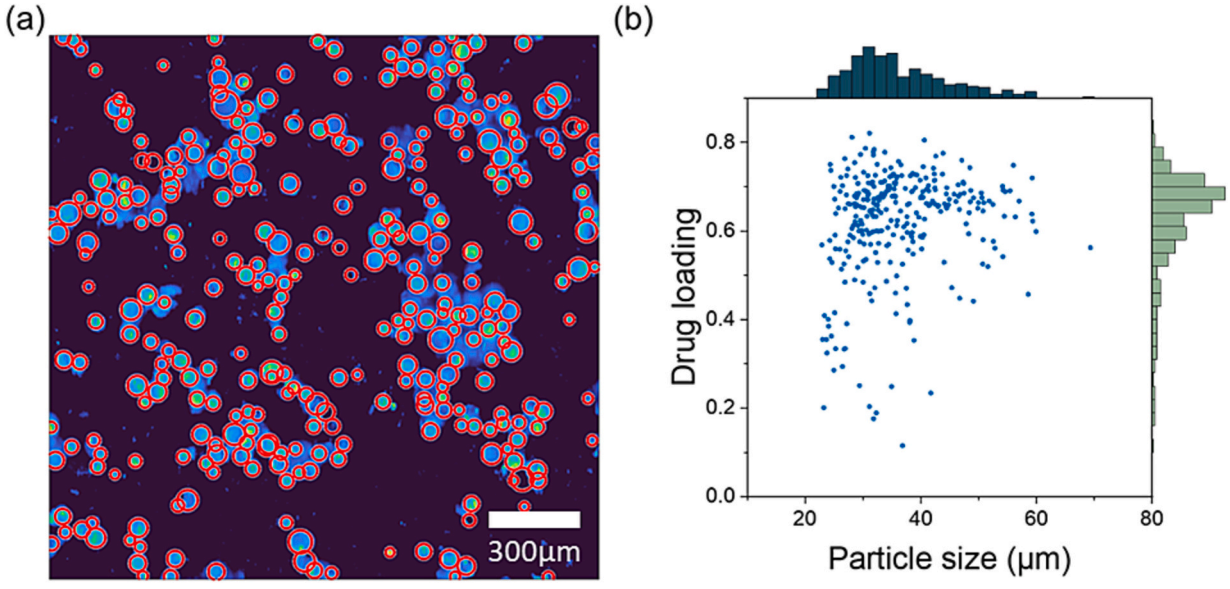

Fig. 5. (a) Raman intensity map of the MPA distribution in representative MPA microspheres. The red circles are the round target recognition of the individual microsphere. (b) Particle-based drug loading *vs.* particle size of the microspheres. The distribution of particle size and drug loading of the microspheres were plotted on the top and right of the figure, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

machine learning model or the post-processing of the API classification results. The amorphous solid dispersion (ASD) with 20% *w*/w drug loading of MPA was prepared using the solvent evaporation method. The initial amorphous spectrum acquisition was performed by spectrum subtraction from the prepared MPA ASD. To verify the crystal form and/or crystallinity, the XRPD spectra, and DSC were collected for the drug MPA, polymer for ASD (PVA VA64), PLGA (RG 752H), the prepared ASD, and MPA microspheres.

In the XRPD profiles (Fig. 6a), both the pure drug and the MPA microspheres M6 exhibited high crystallinity. While the polymers PVA VA64, PLGA, and ASD show no noticeable peak, indicating a complete amorphous state. These results were consistent with the DSC curves (Fig. 6b) since the pure polymers and ASD show no drug melting peak, while the melting peak was observed in the microsphere formulation M6. The amorphous MPA Raman spectrum was obtained by subtraction between the MPA ASD and pure PVA VA64. Fig. 6c shows the crystal and the processed amorphous spectrum of MPA. Within the spectrum range, there were three major difference regions labeled by green, orange, and blue squares. Among the three regions, the orange and blue squared regions were not intrinsic amorphous absorption but interference as a result of the strong signal overlapping with the VA64. Therefore, the green squared region was selected as the amorphous characteristic peaks of MPA. In Fig. 6d, one of the amorphous pixel spectra from the microsphere was chosen for direct comparison with the reference spectrum of the crystal and amorphous MPA. There is an obvious agreement between the amorphous spectrum of ASD and the amorphous pixel spectra of microsphere. Applying the spectrum classification based on amorphous characteristic peaks, 2024 out of 96,245 MPA pixels were recognized as the amorphous state (2.10%).

3.5. *Content uniformity and heterogeneity of the Risperidone microspheres*

Commercially available Risperidone microsphere formulation (Risperdal Consta) contains 38.1% of the drug content [36]. A series of Risperidone microspheres with different drug loading were prepared to explore the maximum drug loading, content uniformity and heterogeneity of the formulations.

The homogeneity of the microsphere may impact the release profile and clinical performance. Monitoring the homogeneity differences between the microsphere formulations with different drug loading can help guide the optimization of formulation and processing parameters. Previous reports have investigated the microstructure and inter-sphere homogeneity of PLGA-based microspheres using XRM and FIB-SEM [1,2]. In Raman imaging, the homogeneity of the sample is revealed by the Raman intensity map of risperidone in Fig. 7a. The particle-based drug heterogeneity was illustrated by the standard deviation of the chemical image of risperidone. The particle identification algorithm will

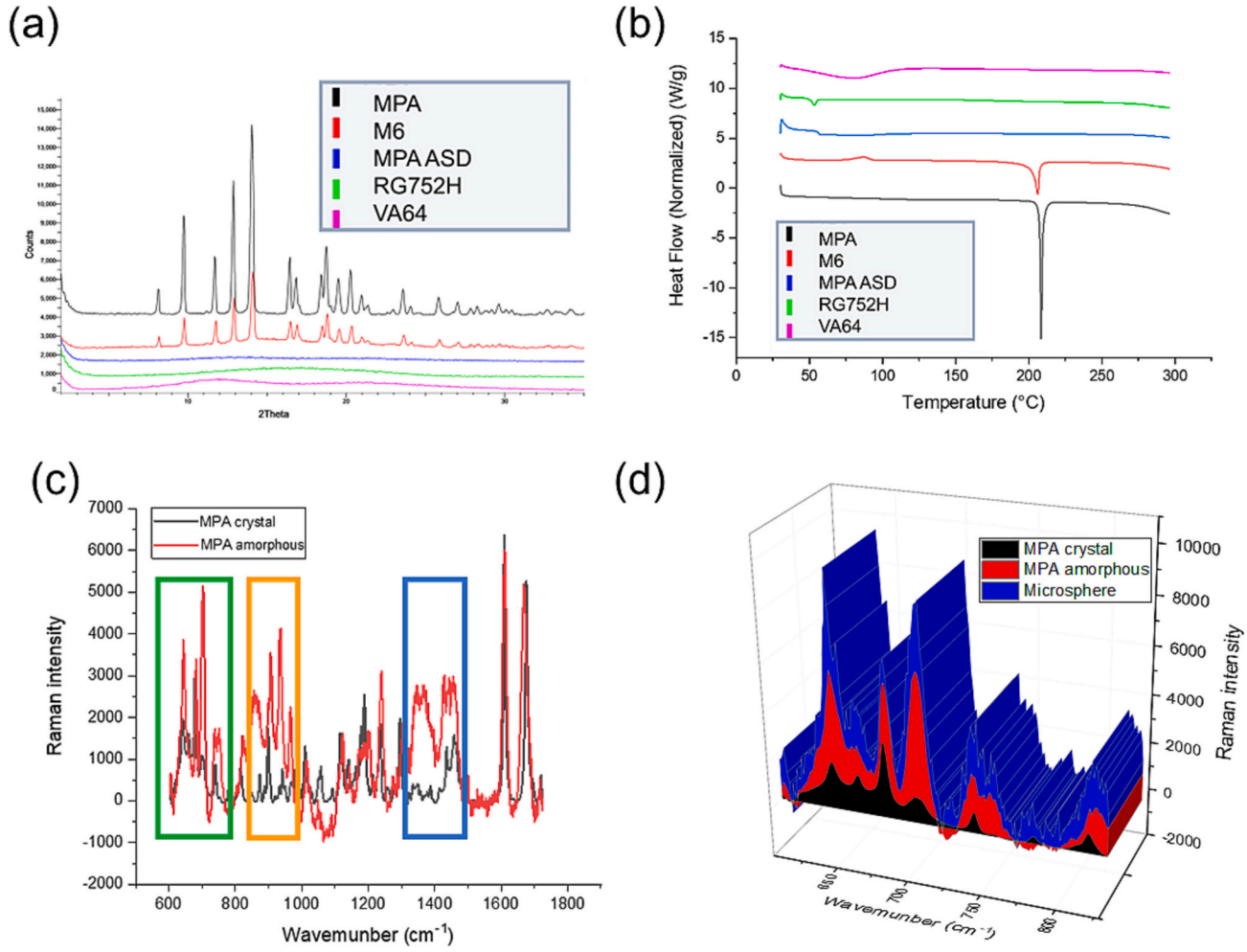

Fig. 6. (a) XRPD profiles of MPA API, MPA microspheres M6, MPA/PVA VA64 amorphous solid dispersion with 20% *w*/w drug loading, PLGA polymer (RG752H), and PVP VA64. The MPA and microspheres demonstrated strong crystal signals while the rest of the samples were highly amorphous. (b) DSC profiles of the samples mentioned in (a). (c) The Raman spectra of the crystal and amorphous MPA. The three major difference regions are labeled by green, orange, and blue squares. (d) The spectrum comparison between the amorphous spectrum in microsphere M6, MPA crystal, and amorphous reference spectrum. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

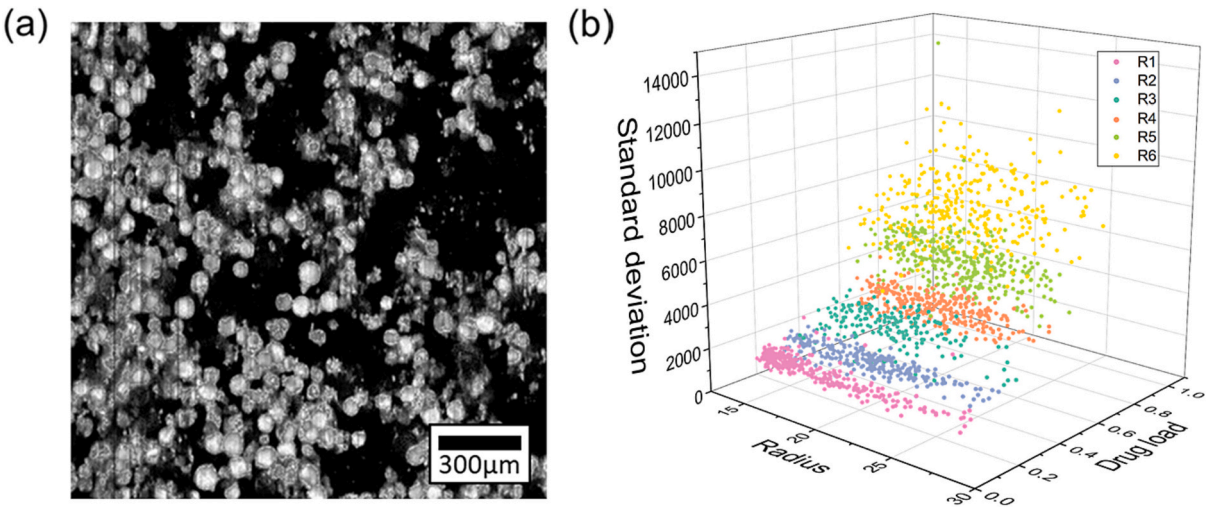

Fig. 7. (a) Risperidone confocal Raman intensity map inside PLGA microsphere. (b) Particle-based statistical analysis for different batches of the Risperidone microspheres.

crop the image of each microsphere and perform a 2D array standard deviation. Each microsphere is assigned with its drug loading, particle size, and heterogeneity (standard deviation). All the Risperidone microspheres were then plotted in a 3-axis volume in Fig. 7b.

As shown in Fig. 7b, formulations R1, R2, R3, and R4 showed low standard deviation, whereas formulations R5 and R6 showed high standard deviation, indicating higher heterogeneity in these formulations. It is worth noting that when the drug loading increases over 40% (*w*/w), the particle size distributions remain similar, however, the variation in drug loading and the drug heterogeneity within a single microsphere increased dramatically. Meanwhile, significant drug crystallization and heterogeneous distribution inside microspheres were observed under PLM for the formulation R6 with the highest drug loading (Fig. 4). In view of this, it is recommended that the maximum drug loading of the final Risperidone formulation should not exceed 40% w/w during the formulation development. This could be one of the reasons why the commercially available Risperidone microspheres have a drug loading close to 40% (*i.e.*, 38.1% (w/w)).

4. Conclusion

For the first time, machine-learning enhanced hyper-spectrum imaging analysis was developed and applied to PLGA microspheres through particle-based physical and chemical classification of the individual particles. The chemical information can be collected simultaneously and automatically using a Raman focus tracking imaging system. Coupling the Raman imaging with machine learning algorithm, the key characteristics of the microspheres (*e.g.*, drug loading, particle size and distribution, amorphous content, heterogeneity of each microsphere) can be determined with simplicity and high accuracy, without extensive separate characterization and analysis of each attribute. The proposed measurement workflow requires no sample preparation pre-measurement, low sample volume (~1 mg), and noninvasive detection, which is especially significant in early formulation screening of microspheres. With automatic data processing, the multidimensional visualization of the result can extensively accelerate early-stage formulation development. In conclusion, the developed novel approach remarkably reduces the complexity of the microsphere screening and enhances the accuracy and reliability of decision-making.

CRediT authorship contribution statement

Minghe Li: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ruifeng Wang:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Quanying Bao:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Data availability

No data was used for the research described in the article.

Acknowledgments

The authors are grateful for the assistance and support from Dr. Yin-Chao Tseng, Ms. Michelle Raikes, Dr. Scott Pennino, Dr. Qi Jiang, Ms. Suja Williams, Ms. Lifen Wu, and Dr. Laibin Luo from the Material and Analytical Sciences Department, as well as Dr. Fangfang Chen from DMPK Department, at Boehringer Ingelheim Pharmaceuticals, Inc., Ridgefield, USA.

References

[1] A.G. Clark, R. Wang, J. Lomeo, Y. Wang, A. Zhu, M. Shen, Q. Bao, D.J. Burgess, B. Qin, S. Zhang, Investigating structural attributes of drug encapsulated microspheres with quantitative X-ray imaging, J. Control. Release 358 (2023) 626–635.
[2] A.G. Clark, R. Wang, Y. Qin, Y. Wang, A. Zhu, J. Lomeo, Q. Bao, D.J. Burgess, J. Chen, B. Qin, Assessing microstructural critical quality attributes in PLGA microspheres by FIB-SEM analytics, J. Control. Release 349 (2022) 580–591.
[3] B. Wan, Q. Bao, D.J. Burgess, In vitro-in vivo correlation of PLGA microspheres: effect of polymer source variation and temperature, J. Control. Release 347 (2022) 347–355.
[4] B. Wan, Q. Bao, R. Wang, D.J. Burgess, Polymer source affects in vitro-in vivo correlation of leuprolide acetate PLGA microspheres, Int. J. Pharm. 625 (2022) 122032.
[5] R. Wang, Q. Bao, A.G. Clark, Y. Wang, S. Zhang, D.J. Burgess, Characterization and in vitro release of minocycline hydrochloride microspheres prepared via coacervation, Int. J. Pharm. 628 (2022) 122292.
[6] FDA organge book, Url: https://www.fda.gov/drugs/drug-approvals-and-databases/approved-drug-products-therapeutic-equivalence-evaluations-orange-book, 2023. accessed:09.22.2023.
[7] M.D. Duncan, J. Reintjes, T. Manuccia, Scanning coherent anti-stokes Raman microscope, Opt. Lett. 7 (1982) 350–352.
[8] C.M. McGoverin, T. Rades, K.C. Gordon, Recent pharmaceutical applications of Raman and terahertz spectroscopies, J. Pharm. Sci. 97 (2008) 4598–4621.
[9] J.-X. Cheng, X.S. Xie, Coherent Raman Scattering Microscopy, CRC press, 2016.
[10] R. Swain, M. Stevens, Raman microspectroscopy for non-invasive biochemical analysis of single cells, Biochem. Soc. Trans. 35 (2007) 544–549.
[11] W.-T. Wang, H. Zhang, Y. Yuan, Y. Guo, S.-X. He, Research progress of Raman spectroscopy in drug analysis, AAPS PharmSciTech 19 (2018) 2921–2928.

[12] V. Calvino-Casilda, M. Bañares, E. LozanoDiz, Real-time Raman monitoring during coumarins synthesis via Pechmann condensation: a tool for controlling the preparation of pharmaceuticals, Catal. Today 155 (2010) 279–281.
[13] N.E. Leadbeater, J.R. Schmink, Use of Raman spectroscopy as a tool for in situ monitoring of microwave-promoted reactions, Nat. Protoc. 3 (2008) 1–7.
[14] P.J. Larkin, M. Dabros, B. Sarsfield, E. Chan, J.T. Carriere, B.C. Smith, Polymorph characterization of active pharmaceutical ingredients (APIs) using low-frequency Raman spectroscopy, Appl. Spectrosc. 68 (2014) 758–776.
[15] M.C. Hennigan, A.G. Ryder, Quantitative polymorph contaminant analysis in tablets using Raman and near infra-red spectroscopies, J. Pharm. Biomed. Anal. 72 (2013) 163–171.
[16] N. Chieng, T. Rades, J. Aaltonen, An overview of recent studies on the analysis of pharmaceutical polymorphs, J. Pharm. Biomed. Anal. 55 (2011) 618–644.
[17] Y. Roggo, K. Degardin, P. Margot, Identification of pharmaceutical tablets by Raman spectroscopy and chemometrics, Talanta 81 (2010) 988–995.
[18] E. Widjaja, R.K.H. Seah, Application of Raman microscopy and band-target entropy minimization to identify minor components in model pharmaceutical tablets, J. Pharm. Biomed. Anal. 46 (2008) 274–281.
[19] T.F. Haefele, K. Paulus, Confocal Raman Microscopy in Pharmaceutical Development, Confocal Raman Microscopy, Springer, 2010, pp. 165–202.
[20] F.G. Vogt, M. Strohmeier, Confocal UV and resonance Raman microscopic imaging of pharmaceutical products, Mol. Pharm. 10 (2013) 4216–4228.
[21] N.M. Ralbovsky, I.K. Lednev, Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning, Chem. Soc. Rev. 49 (2020) 7428–7453.
[22] F. Lussier, V. Thibault, B. Charron, G.Q. Wallace, J.-F. Masson, Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering, TrAC Trends Anal. Chem. 124 (2020) 115796.
[23] W. Hu, S. Ye, Y. Zhang, T. Li, G. Zhang, Y. Luo, S. Mukamel, J. Jiang, Machine learning protocol for surface-enhanced raman spectroscopy, The journal of physical chemistry letters 10 (2019) 6026–6031.
[24] R. Burbidge, M. Trotter, B. Buxton, S. Holden, Drug design by machine learning: support vector machines for pharmaceutical data analysis, Comput. Chem. 26 (2001) 5–14.
[25] S. Kolluri, J. Lin, R. Liu, Y. Zhang, W. Zhang, Machine learning and artificial intelligence in pharmaceutical research and development: a review, AAPS J. 24 (2022) 1–10.
[26] K. Abbas, M. Afaq, T. Ahmed Khan, W.-C. Song, A blockchain and machine learning-based drug supply chain management and recommendation system for smart pharmaceutical industry, Electronics 9 (2020) 852.
[27] V. Dubey, T.R. Saini, Formulation development and pharmacokinetic studies of long acting in situ depot injection of risperidone, Braz. J. Pharm. Sci. 58 (2022) e18809.
[28] Q. Bao, Y. Zou, Y. Wang, S. Choi, D.J. Burgess, Impact of formulation parameters on in vitro release from long-acting injectable suspensions, AAPS J. 23 (2021) 1–11.
[29] B. Wan, Q. Bao, Y. Zou, Y. Wang, D.J. Burgess, Effect of polymer source variation on the properties and performance of risperidone microspheres, Int. J. Pharm. 610 (2021) 121265.
[30] M. Kohno, J.V. Andhariya, B. Wan, Q. Bao, S. Rothstein, M. Hezel, Y. Wang, D. J. Burgess, The effect of PLGA molecular weight differences on risperidone release from microspheres, Int. J. Pharm. 582 (2020) 119339.
[31] J. Shen, S. Choi, W. Qu, Y. Wang, D.J. Burgess, In vitro-in vivo correlation of parenteral risperidone polymeric microspheres, J. Control. Release 218 (2015) 2–12.
[32] A. Rawat, E. Stippler, V.P. Shah, D.J. Burgess, Validation of USP apparatus 4 method for microsphere in vitro release testing using Risperdal® Consta®, Int. J. Pharm. 420 (2011) 198–205.
[33] D.A. Long, Raman Spectroscopy, New York 1, 1977.
[34] K.N.G. Hoffmeister, Development and Application of High-Speed Raman/Rayleigh Scattering in Turbulent Nonpremixed Flames, The Ohio State University, 2015.
[35] J. Shen, D.J. Burgess, Accelerated in-vitro release testing methods for extended-release parenteral dosage forms, J. Pharm. Pharmacol. 64 (2012) 986–996.
[36] Risperdal Consta Product Monograph. https://pdf.hres.ca/dpd_pm/00040784.PDF, 2024.