



 Cite this: *RSC Adv.*, 2018, 8, 23348

Raman spectroscopy coupled with principal component analysis to quantitatively analyze four crystallographic phases of explosive CL-20†

 Xuan He, ^{*,a} Yu Liu,^a Shiliang Huang,^a Yi Liu,^a Xuemei Pu^{*,b} and Tao Xu^{*,a}

The polymorphic quantitative analysis of explosives is very important for national defense and security inspection. However, conventional analytical methods are inaccurate and time-consuming because of the complexity of the polymorphic explosive samples. In this paper, we established a new method of polymorphic quantitative determination in a simple, sensitive, and accurate way. High quality spectra of the four phases of the explosive CL-20 were obtained using a compact Raman spectrometer, and QM calculations were performed to confirm the tentative assignment of the most predominant Raman peaks. Principal component analysis (PCA) of the data was performed to understand the factors affecting the spectral variation across the entire Raman region of the four phases of CL-20 and to calculate the characteristic Raman shift region. In addition, different characteristic peaks were selected according to the PCA and QM calculation results, and a new method for the quantitative determination of polymorphic impurities in ϵ -CL-20 was also set up. The detection level for the polymorphic impurities was determined to be below 1%, and the standard deviation was less than $\pm 0.5\%$. This new method is of significant importance for the quality control of synthesis and production not only in explosives, but also in pharmaceuticals, agrochemicals, and optics industries.

Received 13th March 2018

Accepted 9th June 2018

DOI: 10.1039/c8ra02189a

rsc.li/rsc-advances

Introduction

In materials science, because of different molecular arrangements in the crystal lattice, polymorphism is considered as the ability of a compound to crystallize in more than one form.¹ Different polymorphs of drugs, explosives, agrochemicals, pigments, and dyestuffs can exhibit dissimilar physical and chemical properties, many of which may translate into differences in actual applications.^{2–6} The ability to detect the presence of different polymorphs in a sample, assay their relative amounts, and ultimately obtain the desired form in very high yield is of crucial importance in chemical engineering and technology.^{7–10}

Raman scattering is a powerful method to investigate polymorphs due to its remarkable sensitivity to the crystalline structure of materials.^{11–15} The ability to determine the simultaneous presence of multiple polymorphs in the same sample by distinguishing their Raman spectra is a challenge in polymorph characterization because the vibration frequency of Raman peaks is very weak, complex, and often hard to

distinguish. Therefore, we need to introduce theoretical calculations for a more accurate spectral analysis.

Principal component analysis (PCA) is a classic multivariate technique that analyses a data table where observations are described by several inter-correlated quantitative dependent variables.^{16,17} PCA of the data is performed to understand the factors affecting the spectral variation across the samples, and it is calculated from the covariance matrix of the original data set. PCA is an axis rotation technique that aligns a new set of axes, called principal components (PCs), with the maximal directions of variance within a dataset. Thus, this method creates three new matrices containing the scores, the loadings, and the residuals.^{17,18} The score plot shows whether the samples or a group of samples are different. The corresponding loading plot shows the difference in the Raman spectra and different polymorphs of the samples. Therefore, using the principal component analysis method, the purpose is to interpret the complex Raman spectra of different polymorph compounds by revealing differences between the samples (expressed as so-called “scores”) and relating them to differences in the variables (called “loadings”) defining a sample.^{19–22} Successful applications of PCA to understand the factors affecting spectral variation have been reported.

Fig. 1a shows one of the most recently developed energetic materials: the caged nitramine 2,4,6,8,10,12-hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane (CL-20). There are four different polymorphic phases at ambient pressure and

^aInstitute of Chemical Materials, China Academy of Engineering Physics, Mianyang 621900, China. E-mail: xuan.hellen@gmail.com; xuan.hellen@caep.cn

^bCollege of Chemistry, Sichuan University, Chengdu 610064, People's Republic of China

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8ra02189a



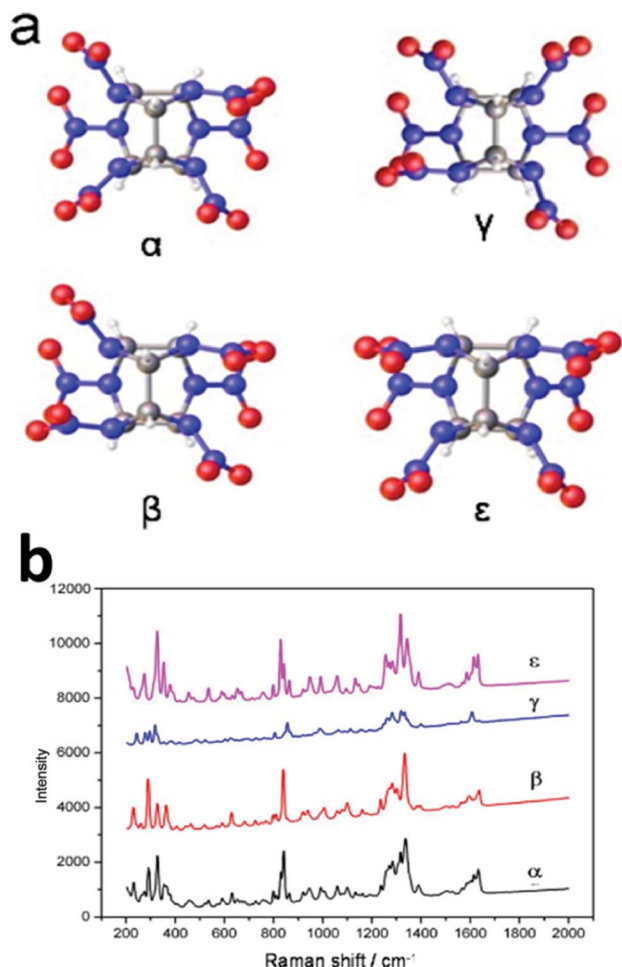


Fig. 1 (a) Structure of the caged nitramine 2,4,6,8,10,12-hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane (CL-20). (b) Raman spectra of the four different phases of CL-20 (α , β , γ , and ϵ).

temperature: α , β , γ , and ϵ . The ϵ -phase is the most thermodynamically stable and the highest density energetic material.^{23–25} In this study, we characterized the four phases of the CL-20 explosive by Raman spectroscopy using the PCA method to draw the main spectroscopic information. A tentative assignment of the most predominant Raman peaks were recorded with the aid of QM calculations at the B3LYP/6-311G(d) level. Different characteristic peaks were selected according the PCA and QM calculation results, and a new method for the quantitative determination of polymorphic impurities in ϵ -CL-20 was also developed. The detection level for polymorphic impurities was determined to be below 1%. This new method is of significant importance for the quality control of synthesis and production not only in explosives, but also in pharmaceuticals, agrochemicals, pigments, dyestuffs, food, energy storage, biominerals, and nonlinear and optics industries.

Experimental

Apparatus and software

The crystal structure of the products was studied using a Bruker D8 X-ray diffractometer with Cu K α irradiation at $\lambda = 1.5406 \text{ \AA}$.

The Raman test conditions were the following: laser wavelength, 532 nm; laser energy, 5 mW; wavenumber range, 3500–50 cm^{-1} ; resolution, 32-times the 4 cm^{-1} -background acquisition; number of samples, 32. DXR Smart Raman Nicolet spectrometer was used for the measurements. To check the sample homogeneity, all analyses of the polymorph mixtures were performed in triplicate. The spectra shown are averages of the triplicate runs.

All data obtained from the experiments were gathered in a data matrix using Microsoft Office Excel (version 2010) and transferred to the MATLAB software. All calculations were conducted using MATLAB (version 2013 a).

Sample preparation

PCA calculations. CL-20 samples (3 mg) were packed into a sample cup for analysis. The Raman spectrogram of the α , β , γ , and ϵ -phase of CL-20 were utilized to construct the calibration set (Fig. 1b). The calibration set consisted of 60 samples, with 15 samples of each phase.

Analysis of polymorphic purity. In brief, 0.1 g of pure β - and ϵ -polymorphs of CL-20 were weighed, mixed, and ground. Two groups of samples were prepared: the first group: β -100%, β -10% + ϵ -, β -20% + ϵ -, β -30% + ϵ -, β -40% + ϵ -, β -50% + ϵ -, β -60% + ϵ -, β -70% + ϵ -, β -80% + ϵ -, β -90% + ϵ -, ϵ -100%, the second group: β -1% + ϵ -, β -3% + ϵ -, β -5% + ϵ -. Another two-phase sample was prepared in the same way.

Results and discussion

The calibration set: construction of PCA models

Experimental design. The XRD and Raman spectra of the four phases of CL-20 (α , β , γ , and ϵ) are shown in Fig. S1 in the ESI† and in Fig. 1b. The Raman spectra of all four stable phases of CL-20 were measured from 50 to 3500 cm^{-1} . The measured Raman spectra usually contain two types of noise: one is the additive noise from external conditions, and the other is the background noise from auto-fluorescence. These noises should be carefully treated because they have a strong effect on the performance of the analysis system. Prior to PCA, all spectra were smoothed and normalized to an equal average intensity value, which was then subtracted from each spectrum. After the noise reduction, PCA was performed.

The 60 spectra (15 spectra for each phase) were divided into four groups and then, they were added to the algorithm program. By taking advantage of PCA, we can clearly see the most significant differences in the Raman spectra of the four crystal types of CL-20, thus providing strong support for the establishment of the standard method of crystal quantitative analysis. Moreover, the final goal of this study was to develop a novel strategy to identify the different phases of the CL-20 explosive using Raman spectroscopy coupled with PCA. The elemental composition of CL-20 with different phases can be visualized by two- or three-dimensional cluster analysis to statistically reduce the dimensionality of the elemental composition data to a smaller set of theoretically meaningful component variables. Visual classification based on the first three principle components for the four binary classifiers was



carried out for the discriminate analysis of the feature vectors obtained from the Raman spectral data.

The training data (Raman spectra from 50 to 3500 cm^{-1}) was included in the corresponding PCA except for the spectra from an individual cell dish. Those Raman spectra usually contain a large number of signals that add complexity to the classification problem. Therefore, the set of spectra was arranged in such a way that each row contained the mean-centered spectrum of a single sample, and the set was polished to remove the less useful regions. A 120×514 (row \times column) matrix was established. Subsequently, we obtained a score plot analysis of the principal components (PCs) of the input matrix. As shown in the calculation process, the input matrix takes into consideration up to six PCs, as the first PC describes only 51.01% of the data variation, confirming the complexity of the system.

Next, examination of the cumulative variance plot revealed that three components reflected 89.79% of the system variation and six components accounted for 91% of the variance of the system. In Fig. 2, the MAP classification shows that the separation between the different phases of CL-20 was not very clear. According to these calculation results, we should classify the spectral regions and select the specific regions by the PCA method to obtain the 3D feature classification map of the region.

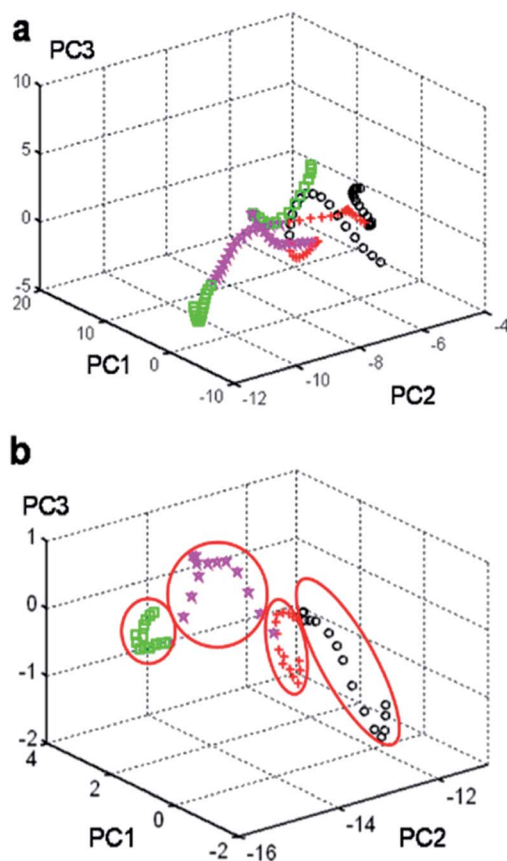


Fig. 2 (a) MAP classification showing that the input feature vectors are mixed in the group region from 50 to 3500 cm^{-1} for the four phases of CL-20. (b) MAP classification showing that the input feature vectors are well separated in the group region from 800 to 1000 cm^{-1} for the four phases of CL-20. (\square : α -CL-20; \star : β -CL-20; $+$: γ -CL-20; \circ : ϵ -CL-20).

Selection of the number of factors and optimized calculated models

Next, a tentative assignment of the peaks was completed with the assistance of quantum mechanics (QM) calculations, and the structure was optimized at the B3LYP-6-311G++(d,p) level of theory using Gaussian 03 (see Fig. 1a and Table S1†). Interestingly, the peaks were unique to the different polymorphs. In the full spectra (50 to 3500 cm^{-1}), there were several characteristic peaks for the various polymorphs. It was evident that it is possible to determine the polymorphic identity of an unknown sample of CL-20 by detailed examination. It is worth mentioning that there are strong peaks for all four isomers, probably ascribed to the ring stretching in the cage and to NO_2 scissoring in the area around 800 cm^{-1} (Fig. 3a).²⁶ There are characteristic peaks for three of the four isomers in this area. The bands at 848 cm^{-1} , 838 cm^{-1} , and 828 cm^{-1} correspond to the characteristic bands of α -, β -, and ϵ -phases, respectively. Another area with strong bands ranges from 260 to 300 cm^{-1} , in which in the area from 280 to 290 cm^{-1} , there are peaks corresponding to all polymorphs except for the ϵ -phase.²⁷ As a result, the 800–850 cm^{-1} and 250–300 cm^{-1} areas are the key regions that require further the analysis.

Next, a series of 60 independent samples with Raman spectra from 50 to 3500 cm^{-1} was split into 7 groups (50–200 cm^{-1} , 200–400 cm^{-1} , 400–800 cm^{-1} , 800–1000 cm^{-1} , 1000–1500 cm^{-1} , 1500–2500 cm^{-1} , and 2500–3500 cm^{-1}). After calculations and based on the MAP classification results, the input matrix indicated that up to six PCs should be taken into consideration as

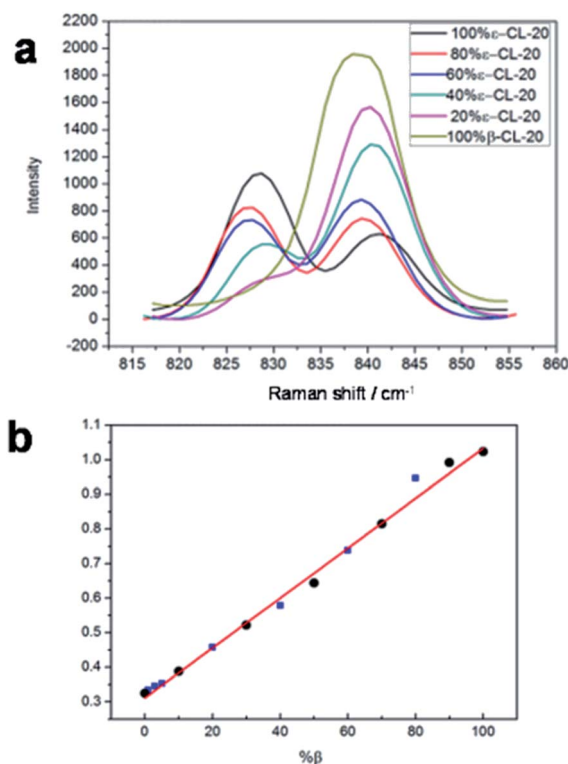


Fig. 3 (a) Raman spectra of CL-20 (mixtures of β and ϵ) in the 810–870 cm^{-1} spectral region; (b) linear relationship between the Raman intensity and the logarithm of the mixtures of β - and ϵ -CL-20.



the first PC described only 76.05% of the data variation. Moreover, examination of the cumulative variance plot revealed that three components reflected 97.07% of the system variation, but six components accounted for 99.86% of the variance of the system. The MAP classification clearly showed that the input feature vectors were mixed and could not be separated in the group region from 50–3500 cm^{-1} (Fig. 2a). However, they could be well separated using this feature space in the group region from 800–1000 cm^{-1} (Fig. 2b). The other calculated results are shown in the ESI† Fig. S3–S8.† The position of each phase of the CL-20 compound was reported in the three-dimensional space for the three principal components (PC1, PC2, and PC3). Consequently, the area from 800–1000 cm^{-1} was the feature region that displayed the maximal difference. Moreover, the quantum mechanics (QM) calculation (tentative assignment of peaks) results showed strong peaks in all four isomers, probably relating to the ring stretching in the cage in the area around 800 \pm 100 cm^{-1} (Table S1†). Analysis of the Raman spectra of all four polymorphs prompted us to develop a new method for the polymorphic purity analysis of ϵ -CL-20.

New method for polymorphic detection

The linear relationship between the ϵ -CL-20 content and that of the other phases of CL-20 was established based on the relative peak height of the ring stretching in the cage and the NO_2 scissoring. The following equation is obtained:

$$X = a + kY$$

where X is the mixture content of α -, β - or γ - in ϵ -CL-20, a is a constant, k is the coefficient, and Y is the peak of the two strong rates of change, which is determined as

$$Y = I_{\epsilon} / I_{\epsilon} + I_C$$

where I is the Raman intensity of the characteristic peak of each phase of CL-20, C is α -, β - or γ -CL-20. Furthermore, the 810–870 cm^{-1} region of the Raman spectra was analyzed. The overlay of the 810–870 cm^{-1} expansion of the spectra generated is displayed in Fig. 3a. A series of experiments were performed on mixtures of β - and ϵ -CL-20, which were varied from 100% β - to 100% ϵ - (β -100%, β -10% + ϵ -, β -20% + ϵ -, β -30% + ϵ -, β -40% + ϵ -, β -50% + ϵ -, β -60% + ϵ -, β -70% + ϵ -, β -80% + ϵ -, β -90% + ϵ -, and ϵ -100%). Fig. 3b illustrates the linear relationship between the Raman intensity and the logarithm of the mixtures of β - and ϵ -CL-20 in the range of β -100%, β -10% + ϵ -, β -30% + ϵ -, β -50% + ϵ -, β -70% + ϵ -, β -90% + ϵ -, and ϵ -100%. The peak of the two strong rates of change, Y , was $I_{838}/I_{838} + I_{828}$, where I_{828} is the intensity of the peak at 828 cm^{-1} of ϵ -CL-20 and I_{838} is the intensity of the peak at 838 cm^{-1} of β -CL-20. The overlay of the 800 \pm 100 cm^{-1} expansion of the generated spectra is displayed in Fig. 3b. From the linear regression equation $y = 7.442 - 0.738x$, the square of the correlation coefficient was calculated to be 0.9974. The standard deviation was 0.0158, $N = 3$. Then, more samples of the mixtures of β/ϵ in β -20% + ϵ -, β -40% + ϵ -, β -60% + ϵ -, and β -80% + ϵ - were detected and calculated into the linear regression equation. The standard deviation was less than 3%.

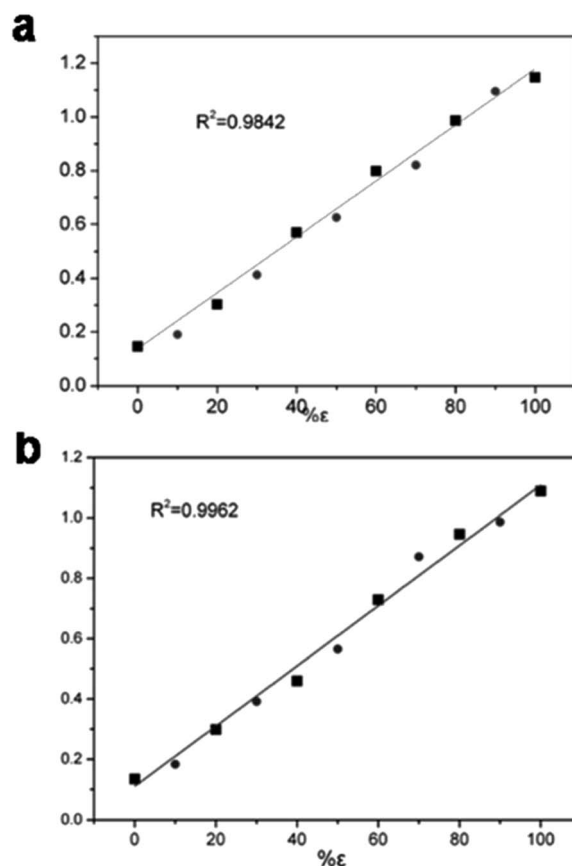


Fig. 4 (a) Linear relationship between the Raman intensity and the logarithm ($I_{838}/I_{838} + I_{828}$) of the mixtures of α - and ϵ -CL-20. (b) Linear relationship between the Raman intensity and the logarithm ($I_{838}/I_{838} + I_{828}$) of the mixtures of γ - and ϵ -CL-20.

Applications to analysis of polymorphic purity

It is of significance to check the polymorphic purity of the ϵ -phase for very small amounts of the mixture phase.^{28,29} The real samples were made by varying the amount of β -phase in the ϵ -phase from 1, 3, to 5%. It was possible to manually detect 1% of the β -phase in the ϵ -phase, and the standard deviation was less than $\pm 0.5\%$ (Fig. 3b). To further ascertain that this quantitative analysis method had good applicability, we prepared other two-phase mixtures (α and γ -phase) in ϵ -phase CL-20 with a similar method that is detailed in the ESI† Fig. S9 and S10.† The linear relationship between the Raman intensity and the logarithm of the mixtures of α - and ϵ -CL-20 and the mixtures γ - and ϵ -CL-20 are shown in Fig. 4, Table S3 and S4;† the standard deviation was less than $\pm 2.0\%$ and $\pm 2.5\%$, respectively. Therefore, Raman spectroscopy and PCA calculation methods could be used for the quantitative analysis of low-content crystalline samples in ϵ -CL-20. The results suggested that this methodology is a promising and useful alternative for classifying and assigning the polymorphic identity of other explosives in a complex system.

Conclusions

In summary, a simple, sensitive, and accurate method for the polymorphic quantitative analysis of four different phases of



the explosive CL-20 was developed. A tentative assignment of the most predominant Raman peaks was made with the aid of QM calculations at the B3LYP/6-311G(d) level. The different characteristic peaks were selected based on the PCA analysis and QM calculation results. A new method for the quantitative determination of polymorphic impurities in ϵ -CL-20 was also established. The detection level for polymorphic impurities was determined to be below 1%, and the standard deviation was less than 0.5%. Other two-phase CL-20 mixtures in the ϵ -phase were also detected. This new method is of significance for quality control of synthesis and production not only in explosives, but also in pharmaceuticals, agrochemicals, pigments, dyestuffs, food, energy storage, bio-minerals, and nonlinear and optics industries.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This study was supported by the National Natural Science Foundation (No. 21502179); the Presidential Foundation of CAEP (Grant No. 201501018); the NSAF (Grant No. U1730127).

Notes and references

- 1 J. Bernstein, *Polymorphism in Molecular Crystals*, International Union of Crystallography, Oxford University Press, Oxford, UK, 2008.
- 2 J. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon Press, Oxford, 2002.
- 3 H. G. Brittain, *Polymorphism in Pharmaceutical Solids*, Marcel Dekker, New York, 1999.
- 4 J. Aaltonen, M. Alleso, S. Mirza, V. Koradia, K. C. Gordon and J. Rantanen, Solid form screening-A review, *Eur. J. Pharm. Biopharm.*, 2009, **71**, 23–37.
- 5 D. J. W. Grant and S. R. Byrn, *Adv. Drug Delivery Rev.*, 2004, **56**, 237–239.
- 6 A. Llinàs and J. M. Goodman, *Drug Discovery Today*, 2008, **13**, 198–210.
- 7 C. Carteret, A. Dandeu, S. Moussaoui, H. Muhr, B. Humbert and E. Plasari, *Cryst. Growth Des.*, 2009, **9**, 807–812.
- 8 N. L. Calvo, T. S. Kaufman and R. M. Maggio, *J. Pharm. Biomed. Anal.*, 2015, **107**, 419–425.
- 9 K. L. A. Chan, O. S. Fleming, S. G. Kazarian, D. Vassou, G. D. Chrysikos and V. Gionis, *J. Raman Spectrosc.*, 2004, **35**, 353–359.
- 10 G. Gouadec and P. Colomban, *Prog. Cryst. Growth Charact. Mater.*, 2007, **53**, 1–56.
- 11 H. Abdi and L. J. Williams, *Principal component analysis*, John Wiley & Sons, Inc., 2010, vol. 2, pp. 433–459.
- 12 A. Taleb, J. Diamond, J. J. McGarvey, J. R. Beattie, C. Toland and P. W. Hamilton, *J. Phys. Chem. B*, 2006, **110**, 19625–19631.
- 13 J. W. Chan, D. S. Taylor, S. M. Lane, T. Zwerdling, J. Tuscano and T. Huser, *Anal. Chem.*, 2008, **80**, 2180–2187.
- 14 P. Donfack, G. Multhoff and A. Materny, *ChemBioChem*, 2011, **12**, 1922–1936.
- 15 P. R. T. Jess, M. Mazilu, K. Dholakia, A. C. Riches and C. S. Herrington, *Int. J. Cancer*, 2008, **124**, 376–380.
- 16 D. J. Graham and D. G. Castner, *Biointerphases*, 2012, **7**, 49–60.
- 17 M. Holzweber, T. Heinrich, V. Kunz, S. Richter, C. H. H. Traulsen, C. A. Schalley and W. E. S. Unger, *Anal. Chem.*, 2014, **86**, 5740–5748.
- 18 S. Muramoto, D. J. Graham, M. S. Wagner, T. G. Lee, D. W. Moon and D. G. Castner, *J. Phys. Chem. C*, 2011, **115**, 24247–24255.
- 19 M. C. Biesinger, P.-Y. Paepegaey, N. S. McIntyre, R. R. Harbottle and N. O. Petersen, *Anal. Chem.*, 2002, **74**, 5711–5716.
- 20 D. J. Graham and B. D. Ratner, *Langmuir*, 2002, **18**, 5861–5868.
- 21 M. S. Wagner and D. G. Castner, *Langmuir*, 2001, **17**, 4649–4660.
- 22 M. S. Wagner, D. J. Graham, B. D. Ratner and D. G. Castner, *Surf. Sci.*, 2004, **570**, 78–97.
- 23 A. T. Nielsen, A. P. Chafin, S. L. Christian, D. W. Moore, M. P. Nadler, R. A. Nissan, D. J. Vanderah, R. D. Gilardi, C. F. George and J. L. Flippen-Anderson, *Tetrahedron*, 1998, **54**, 11793–11812.
- 24 N. V. Latypov, U. Wellmar, P. Goede and A. J. Bellamy, *Org. Process Res. Dev.*, 2000, **4**, 156–158.
- 25 T. P. Russell, P. J. Miller, G. J. Piermarini and S. Block, *J. Phys. Chem.*, 1993, **97**, 1993–1997.
- 26 J. J. Xu, J. Sun, K. Zhou, H. Z. Li and Y. J. Shu, *Chin. J. Energ. Mater.*, 2012, **20**, 248–255.
- 27 P. Goede, N. V. Latypov and H. Östmark, *Propellants, Explos., Pyrotech.*, 2004, **29**, 205–208.
- 28 M. Ghosh, V. Venkatesan, N. Sikder and A. K. Sikder, *Cent. Eur. J. Energ. Mater.*, 2013, **10**, 419–438.
- 29 S. Dumas, J. Y. Gauvrit and P. Lanteri, *Propellants, Explos., Pyrotech.*, 2012, **37**, 230–234.

