Method and system of detecting a target analyte

Alstrøm, Tommy Sonne; Schmidt, Mikkel Nørgaard; Larsen, Jan; Frøhling, Kasper Bayer; Boisen, Anja; Schmidt, Michael Stenbæk

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 March 2016 (24.03.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/041842 A1

(51) International Patent Classification:
G01N 21/65 (2006.01)

(21) International Application Number:
PCT/EP2015/070723

(22) International Filing Date:
10 September 2015 (10.09.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14185146.9 17 September 2014 (17.09.2014) EP

(71) Applicant: **DANMARKS TEKNISKE UNIVERSITET** [DK/DK]; Anker Engelunds Vej 101 A, 2800 Kgs. Lyngby (DK).

(72) Inventors: **ALSTRØM, Tommy Sonne**; Bjødstrupvej 7E, 2610 Rødovre (DK). **SCHMIDT, Mikkel Nørgaard**; Store Mølle Vej 17, 1.tv., 2300 Copenhagen S (DK). **LARSEN, Jan**; Birkholmsvej 20, 2800 Kgs. Lyngby (DK). **FRØHLING, Kasper Bayer**; Bagsværd Hovedgade 58, st. B, 2880 Bagsværd (DK). **BOISEN, Anja**; Hasselbakken 10, 3460 Birkerød (DK). **SCHMIDT, Michael Stenbæk**; Rubinsteinsvej 25, 2. th, 2450 Copenhagen SV (DK).

(74) Agent: **GUARDIAN IP CONSULTING I/S**; Diplomvej, Building 381, 2800 Kgs. Lyngby (DK).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: METHOD AND SYSTEM OF DETECTING A TARGET ANALYTE

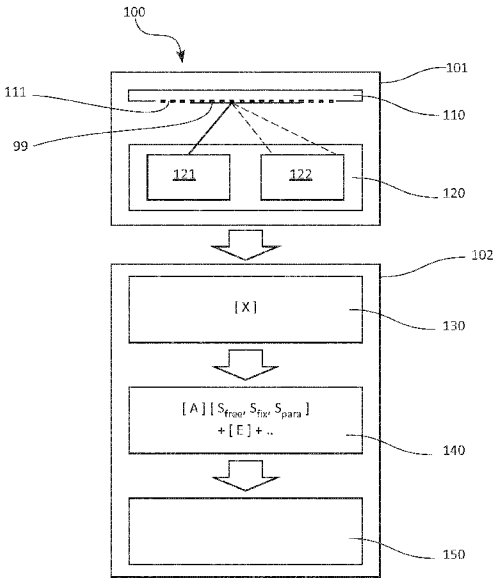

Fig. 1

(57) **Abstract**: The invention relates to a method and system of detecting the presence of one or more specific target analytes in a sample. The method comprises the steps of exposing a SERS-active surface to the sample so as to capture analytes from the sample by adherence to the SERS-active surface; acquiring an ensemble X of Raman/SERS-spectra measured at different locations on the SERS-active surface; for each target analyte, determining a target spectrum representative of that target analyte; modelling the ensemble X of measured Raman/SERS-spectra using a factorization-model, wherein the factorization-model includes a loadings matrix A times a spectral component matrix S, wherein matrix elements of A and S are all real and non-negative; jointly estimating parameters in the factorization-model, the estimated parameters including at least A and S; for each target analyte, selecting amongst the spectral components of S a target component corresponding to the target spectrum of that target analyte; and determining presence of the one or more target analytes at least on the basis of the estimated loadings for the selected target components. During estimation at least one of the one or more target components is populated according to a p re-determined function.

WO 2016/041842 A1

METHOD AND SYSTEM OF DETECTING A TARGET ANALYTE

TECHNICAL FIELD

The present invention relates in one aspect to a method of detecting the presence of a target analyte in a sample by Surface Enhanced Raman Spectroscopy (SERS). In a further aspect the invention relates to a system for detecting the presence of a target analyte in a sample by Surface Enhanced Raman Spectroscopy (SERS). The sample may be in a solid, liquid or gaseous phase.

BACKGROUND OF THE INVENTION

Detection of a specific target analyte, which is present at very low concentrations in a sample, attracts considerable attention. The present invention is concerned with the problem of detecting the presence of one or more specific target analytes in a sample, such as testing for the presence of specific biological and chemical species. The target analytes are specified beforehand. Typically the samples are in a liquid or gaseous phase, but some samples may also be presented in a solid phase, such as in the form of a fine powder with particles in the nanoscale. The result of a detection test is “positive” for a particular target analyte if the presence of that target analyte in the sample can be established. Accordingly, the result is “negative” if the presence of that target analyte in the sample cannot be established. However, this means that the presence of the target analyte in a concentration below the detection limit of a given detection test yields false negative results. Lowering the detection limit for detection test is therefore important for improving the reliability of such tests.

A large variety of techniques have been developed to analyze samples, amongst those optical spectroscopy techniques. When using optical spectroscopy, the energy dependence of the optical response of the sample to an optical stimulus is recorded as a spectrum, plotting e.g. intensity against energy of the optical response, and characteristic features in the spectrum are analyzed to identify constituents of the sample.

Raman spectroscopy is a technique analyzing the inelastic scattering of stimulating photons with molecules where the spectral response exhibits pronounced peaks that are shifted in energy with respect to the stimulus. The energy shifted peaks reflect

events of inelastic interaction between the probing stimulus and characteristic molecular states of molecules in the sample and therefore yield specific information about the sample on a molecular level. However, such events are extremely rare as compared to elastic scattering events. A particularly sensitive Raman spectroscopy technique is Surface Enhanced Raman Spectroscopy (SERS), wherein molecules that are adsorbed to certain SERS-active substrate surfaces exhibit an enhanced Raman scattering signal with an enhancement factor of up to 10E12 or even more, thereby allowing, in principle, for the detection of single molecules. While such enhancement factors may be achieved in certain experimental set-ups, they are in practical applications of detecting a target analyte in a real world sample typically obscured by the unavoidable presence of other substances in the sample.

In the presence of multiple constituents in the same sample, spectral contributions from different constituents may be mixed together in the same spectrum and may therefore be difficult to distinguish from each other. Some attempts to de-mix such spectral information involve the acquisition of large numbers of spectra obtained under more or less systematically varied conditions to vary the contributions from the different constituents. For example, WO 2012/112119 proposes a method for analyzing a liquid sample containing multiple unknown analyte components by Raman spectroscopy, in particular SERS. The method includes forming a layer of analytes on a SERS surface, obtaining Raman spectra of multiple respective pixels, applying a band target entropy minimization (BTEM) algorithm to obtain pure component spectra, and subsequently identifying analytes from the pure component spectra. The proposed method is for extraction of estimated pure component Raman spectra from the experimentally obtained set of mixed Raman spectra. Unknown analytes are identified by comparison of the estimated spectra to known spectra stored in spectral libraries. However, the successful de-mixing and detection relies on the presence of different spectral contributions at various locations of the SERS-active surface. The method is therefore not well-suited for detecting a specific target analyte at very low concentrations.

Typically, the Raman/SERS signal is enhanced through strong localized enhancement of electromagnetic fields observed in so-called hot-spots, such as formed in nano gaps between metal surfaces. However, one concern regarding SERS is the

statistical behavior of the electromagnetic hot spot distribution. It is not only required to find a hot-spot, but there also have to be molecules captured in the hot-spot, which is increasingly unlikely as the concentration of the target molecules decreases. In order to increase the likelihood of capturing a target molecule at a hot-spot of a SERS-active surface, the area of the SERS-active surface that is exposed to the sample may be increased and spectra from the enlarged area may be recorded in a Raman/SERS map. This increases the likelihood that spectra from hot spots containing molecules are recorded. The challenge that arises from this strategy is the compilation of a large amount of irrelevant data diluting the signal of interest from the rare target analyte.

In a scientific article, Li et al., Journal of VLSI Signal Processing (2007), p.83 presents a non-negative matrix factorization algorithm to improve effectiveness of detection of a specific chemical agent in a sample by Raman spectroscopy. The algorithm introduces prior knowledge about the target analyte by means of an orthogonality constraint. The article applies the algorithm to a set of Raman data, but disregards particularities of Raman/SERS-detection on a SERS-active substrate as compared to mere Raman spectroscopy. Consequently, e.g. assessing the validity of the obtained detection result with regard to e.g. false positive detection is not easily possible in this approach.

Therefore, there is a need for an improved method and system of detecting a specific target analyte in a sample at low concentrations.

SUMMARY OF THE INVENTION

A first aspect of the invention relates to a method of detecting the presence of one or more specific target analytes in a sample, the method comprising the steps of

(a) capturing analytes from the sample on a SERS-active surface;

(b) acquiring an ensemble X of Raman/SERS-spectra measured on the SERS-active surface;

(c) for each target analyte, determining a target spectrum representative of that target analyte;

(d) modelling the ensemble X of measured Raman/SERS-spectra using a factorization-model, wherein the factorization-model includes a loadings matrix A times a

spectral component matrix S, wherein matrix elements of A and S are all real and non-negative;

(e) jointly estimating parameters in the factorization-model, the estimated parameters including at least A and S;

(f) for each target analyte, selecting amongst the spectral components of S a target component corresponding to the target spectrum of that target analyte; and

(g) determining presence of the one or more target analytes at least on the basis of the estimated loadings for the selected target components.

In essence, analytes are captured on a SERS-active surface and an ensemble X of Raman/SERS-spectra is acquired from that SERS-active surface. The ensemble X of Raman/SERS-spectra is then processed by a data analysis unit that determines for each target analyte if an instance of positive detection can be established. The data analysis unit comprises a factorization model with parameters that are estimated via statistical inference techniques, such as using a Bayesian frame work, and a classifier that determines a detection result from the model based on the estimated parameters.

The sample containing analytes may be presented in any suitable form including solid, liquid or gas phase. The sample may comprise sub-samples that may have different constituents, different form, and different analyte content. The analytes may be captured on the SERS-active surface in any known manner. Typically, a SERS-active surface is exposed to the sample so as to capture analytes from the sample by adherence to the SERS-active surface. The term “SERS-active surface” is to be understood broadly to also include surface areas located on different SERS-substrates, or even on different SERS-devices, each SERS-device comprising one or more substrates with a surface adapted for SERS-type measurements. Raman spectroscopy measurements are then performed on the SERS-active surface to obtain Raman/SERS-spectra. Most preferably, an ensemble X of Raman/SERS-spectra are measured at different locations on the SERS-active surface. In a particularly advantageous embodiment, at least a subset of the Raman/SERS spectra is measured at contiguous locations in the same portion of the SERS-active surface so as to provide a Raman/SERS spectral map from that portion. In addition to Raman/SERS-spectra from different locations, the ensemble X of Raman/SERS-

spectra may also include measurements taken at the same location on the SERS-active surface, e.g. by repetitive scanning. Advantageously, the SERS-active surface is adapted to enhance the Raman signal in so-called hot-spots. The spectrometer device may be configured for localized acquisition of Raman/SERS-spectral data may be acquired from the SERS-active surface by any known technique, including advanced techniques such as resonant excitation.

A specific analyte to be tested for in the detection is referred to as target analyte. The detection procedure may be directed to testing for the presence of a single target analyte or of multiple target analytes simultaneously. The different target analytes may be present in the same sub-sample or may be present in different sub-samples measured in different measurement campaigns/series and on the SERS-active surface of different SERS-substrates/devices. The method according to the present invention allows for simultaneously analyzing large data sets of Raman/SERS-spectra in order to simultaneously test for the presence of a target analyte. The method even allows for testing for multiple target analytes at the same time and in the presence of unidentified analytes (contaminants), irrespective of whether the target analytes and other analytes are present in the same sub-sample or in different sub-samples, if the spectra are obtained on the same SERS-active substrate or on completely different SERS-devices, as long as the measured data can be assembled to an ensemble X of Raman/SERS-spectra that can be represented in a common matrix.

The ensemble of spectral data is then modelled, wherein modelling includes factorizing the ensemble X of Raman/SERS-spectra as a loadings matrix A times a spectral component matrix S, wherein matrix elements of A and S are all real and non-negative. Real numbers are in the present application denoted by the double-struck letter R. The term “real and non-negative” covers all real numbers including zero and is in the present application denoted by the double-struck letter R with subscript “+”.

$$X = AS \qquad \text{where } X \in \mathbb{R}^{N\times D} \quad A \in \mathbb{R}_{+}^{N\times K} \quad S \in \mathbb{R}_{+}^{K\times D}$$

S is a set of basis vectors s(k,d), wherein each basis vector has D spectrally distributed intensity values (d=1..D) and represents one of K spectral components (k=1..K). A is a set of loading vectors a(n,k) representing the loadings with which the respective S spectral components (k=1..K) contribute to the mixed spectrum for each of the N instances/locations/repetitions/measurements (n=1..N).

Inherently, the factorisation model deviates from the actual measurement data by at least measurement noise. This deviation may be dealt with, e.g. by adding noise contributions as a residual noise matrix E to the product of the loadings matrix A times the spectral components S, wherein matrix elements of E are all real. Since E describes measurement noise, the elements may also assume negative values. Including the measurement noise as additive term, the ensemble X of Raman/SERS-spectra may thus be written as:

$$X = AS + E \qquad \text{where } E \in \mathbb{R}^{N \times D}$$

Measured raw data collected by a detection device configured for the detection of scattered light from the SERS-typically contain a spectral background, which may be dealt with already at acquisition by baseline correction routines commonly known in optical spectroscopy. Alternatively or in addition thereto, any remaining spectral background may be included in the factorization-model by adding background contributions to the product of loadings A times spectral components S as a background signal matrix B, wherein matrix elements of B are all real and non-negative. Including both measurement noise and spectral background as additive terms, the ensemble X of Raman/SERS-spectra may thus be written as:

$$X = AS + B + E \qquad \text{where } B \in \mathbb{R}^{N \times D}$$

The presence of the target analyte may be determined on the basis of all or some of the estimated parameters of the factorization-model. However, one of the advantages of the present invention is that the factorization-model with non-negative elements allows for an intuitive physical interpretation of the model parameters. With regard to determining the presence of the target analyte this means that the basis vectors of S may be interpreted as spectral components of the mixed spectrum,

wherein one or more of these spectral components may be identified as target components, each representing the spectral signature of a particular target analyte. Information on the presence of the target analyte in the sample under investigation may therefore be derived directly from the estimated loadings for the selected target components of S, whereas information on the reliability of detection, the detection limit or the signal-to-noise ratio of the detection may be derived e.g. from the loadings of the target component and the estimated noise matrix E in combination.

More generally, the presence of the target analyte may be determined by a classifier using the factorization model with the estimated parameters as input. The method may thus include feeding estimated factorization model parameters to a classifier, which on the basis of these parameters determines the presence of the one or more target analytes.

Further according to one embodiment of the method, estimation of at least one of the one or more target components is subject to a population rule determining the population of the target component elements with values according to a predetermined function.

Each of the target components of S may be subject to a population rule. The population rule governs the population of a given target component during estimation/inference of the elements of S that make up the target component vector. Thereby, forehand information on the target component in the form of a known spectral reference may in a convenient, flexible and intuitive way be integrated directly in the estimation procedure. The spectral reference may be derived in any suitable way, e.g. from a library of spectral data, a theoretical model, or from a reference measurement for the target analyte. Different target analytes may be subjected to different population rules. By integrating forehand information on the target analyte directly in the factorization model, the sensitivity and reliability of detection of a target analyte at very low concentrations is enhanced significantly as compared to e.g. a blind separation and spectral de-mixing without such forehand information. In case no suitable forehand information about the spectral signature of a given target analyte can be formulated/obtained, a population rule for the corresponding target com-

ponent in S may be omitted. Spectral components for unknown analytes are not subject to any population rule.

Further according to one embodiment of the method, the predetermined function is one of a fixed value target spectrum that is kept fixed during estimation/inference, a parametric description of the target spectrum with target component parameters that are estimated/inferred by the joint estimation; and a probabilistic parametric description of the target spectrum with target component parameters that are estimated / inferred by the joint estimation.

In some embodiments, the values of the target component for a given target analyte are set to the target spectrum selected in step (c) as representative of that target analyte, and the target component is kept fixed during the joint estimation of the parameters in the factorization-model. This strict population rule is particularly useful when a highly reliable reference spectrum is available as target spectrum, e.g. from SERS-measurements on a reference sample containing a high concentration of the target analyte.

In some embodiments, the values of the target component for a given target analyte populated according to a parametric description of the target spectrum selected in step (c) as representative of that target analyte, and the parameters of the parametric target component description are estimated along with the other model parameters during the joint estimation of the parameters in the factorization-model. This parametric population rule is particularly useful e.g. when the target spectrum is available in the form of parametric peak functions describing characteristic spectral features occurring in the target spectrum of the target analyte. Furthermore, a probabilistic parametric function may be useful as a less strict population rule to allow for further flexibility in the description of the target spectrum during estimation, e.g. if the spectral position and/or shape of spectral features in the target spectrum is uncertain. Correspondingly, the parameters of a probabilistic parametric target component description are also estimated along with the remaining parameters of the factorization model.

In some embodiments, the target components of one or more of the target analytes are not subject to any specific population rule and the respective target components are estimated on equal footing with the other spectral components of the spectral component matrix S. After the joint estimation of the parameters of the factorization-model these target components are then selected amongst the estimated spectral components of S by comparison to the respective target spectra determined in step (c), which may be, for example, a reference spectrum from a spectral data library, a theoretical model, or a measurement on a reference sample of the target analyte.

Further according to one embodiment of the method, at least a first target component and a second target component are subject to respective population rules, wherein the predetermined functions determining the population of the first and second target components are different from each other. The population for some or all target components may be the same. However, depending on the available information on the spectral signatures for different target analytes, different population rules may formulated for each of the target components.

For each target component subject to a population rule, one population function type is selected. Target components populated according to different functions may be combined in the same model. Choice of the population rule may e.g. depend on the quality/availability of spectral information about the various target analytes.

Further according to one embodiment of the method, the target spectrum of at least one target analyte in step (c) is determined from reference Raman/SERS-measurements of said at least one target analyte captured on a SERS-active surface. Thereby a target spectrum coming close to the expected experimental conditions of the detection situation is obtained. Preferably, the SERS-active surface for measuring the target spectrum is of the same type as the SERS-active substrate used for the detection measurement, thereby closely matching the experimental conditions of the detection situation.

Further according to one embodiment of the method, the target analyte is captured on the SERS-active surface by exposing the SERS-active surface to a reference sample comprising the target analyte. The reference sample comprises the target

analyte with a sufficiently high concentration so as to obtain a target spectrum with well pronounced spectral features and with a high signal to noise ratio. Advantageously, the concentration of the target analyte should be such as to allow for a clear distinction of the Raman/SERS-signal form the target analyte in the spectra measured on the reference sample. For example, a concentration of target analyte corresponding to a monolayer or more distributed across the surface of a SERS-active substrate on which the reference Raman/SERS measurements are performed is considered a high concentration that is suitable for obtaining a reliable target spectrum.

Further according to one embodiment of the method, the reference Raman/SERS measurements include an ensemble R of reference Raman/SERS-spectra, and the target spectrum is extracted from a factorization-model of the ensemble R. The target spectrum in step (c) may thus extracted from the ensemble R of reference Raman/SERS-spectra by performing steps analogue to steps (a)–(f) above on the reference sample, wherein the reference sample is prepared with a sufficiently high concentration of the target analyte as discussed above. For step (c), the target spectrum is replaced by a pre-determined reference spectrum for the target analyte – e.g. a bulk Raman spectrum of the target analyte found in a spectrum library or a theoretical model of a Raman/SERS-spectrum for the target analyte. The ensemble R may then be modelled as a product of a loadings matrix times a spectral component matrix and the spectral component that comes closest to the pre-determined reference spectrum is selected as the target spectrum for the detection measurement.

Further according to one embodiment of the method, the factorization-model further includes an additive term with a residual noise matrix E, wherein matrix elements of E are all real. The noise term E may be comprise values that are fixed, that are determined by parametric functions, and/or determined by probabilistic parametric functions.

Further according to one embodiment of the method, the factorization-model further includes an additive term with a background matrix B, wherein matrix elements of B are all real and non-negative. The background term B may be comprise values that

are fixed, that are determined by parametric functions, and/or determined by probabilistic parametric functions. In principle, both the noise contributions and the background contributions may be modelled as further spectral components in the spectral component matrix S, thereby contributing additive with appropriate loadings to the total Raman/SERS-signals. One advantage of using a separate additive term for E and B independent of the factorization AS is found in the priors used for governing the estimation/inference of the different model parameters. By keeping the different contributions separate, it is also more straightforward to construct these priors for the different contributions differently and thereby account for the different expectations on the structure/nature/behaviour of the measurement noise and the background signal as compared to the Raman/SERS-data, which are typically sparse.

Further according to one embodiment of the method, the joint estimation is performed in a Bayesian framework. Preferably, the Bayesian framework uses sparse priors, such as exponential priors in a formulation analogue to the model described in Schmidt et al., "Bayesian Non-negative Matrix Factorization" (2009), Lecture notes in Computer Science, pp.540-547. Further preferably, the sparse priors are applied to the loadings matrix A and the spectral components of S that are not governed by a population rule ("S_{free}"). Advantageously, the spectrally broad, typically smooth nature of the background signal is reflected by a smoothness constraint to B, and the measurement noise term E may be denoted as "weakly informative".

Further according to one embodiment of the method, a total number K of spectral components forming the spectral component matrix S is treated as a parameter of the factorization model, which is estimated by the joint estimation, wherein advantageously, the total number is estimated using non-paramtric Bayesian techniques. Another way of determining the total number of K of spectral components to be included in the factorization model is cross-validation where the method is performed using different numbers of K, and the choice of K is cross-validated against an evaluation of achieved result, e.g. by analysing for the different values of K the deviation of the product of A and S with respect to the ensemble of measured data X.

Further according to one embodiment of the method, a detection result for the presence of a given target analyte is determined as positive, if the loading of the target component exceeds a pre-determined threshold.

Further according to one embodiment of the method, a positive detection result for the presence of a given target analyte in one location on the SERS-active surface is validated by regarding loadings of the corresponding target component in neighbouring locations on the SERS-active surface.

Advantageously according to one embodiment of the method, the parameters of the factorization are estimated in a Bayesian framework using a Gibbs-sampler. A suitable Gibbs sampler is found in the above-mentioned document by Schmidt et al.

A second aspect of the invention relates to a system for detecting the presence of one or more specific target analytes in a sample, the system comprising a SERS-active surface, a spectrometer device configured for the acquisition of Raman/SERS-spectral data from the SERS-active surface, and a computer including devices for data storage and data processing, wherein the computer is configured for performing the steps of

(a) storing an ensemble X of measured Raman/SERS-spectra;
(b) modelling the ensemble X of measured Raman/SERS-spectra using a factorization-model, wherein the factorization-model includes a loadings matrix A times a spectral component matrix S, wherein the matrix elements of A and S are all real and non-negative;
(c) jointly estimating parameters in the factorization-model, the estimated parameters including at least A and S
(d) for each target analyte, selecting amongst the spectral components of S a target component corresponding to a pre-determined target spectrum for that target analyte; and
(e) for each target analyte, determining presence of that target analyte at least on the basis of the estimated loadings for the corresponding target component.

Analytes are captured on a SERS-active surface and an ensemble X of Raman/SERS-spectra is acquired from that SERS-active surface. The ensemble X of

Raman/SERS-spectra is stored in a data storage device, and then processed by a data processing unit that determines, for each target analyte, if an instance of positive detection can be established. The data processing unit comprises a factorization model with parameters that are estimated via statistical inference techniques, such as using a Bayesian frame work, and a classifier that determines a detection result from the model based on the estimated parameters.

The spectrometer device includes an excitation light source adapted to excite Raman scattering events on the SERS-active surface, a detection unit for the spatially and spectrally resolved detection of light scattered from the SERS-active surface and for the conversion of the detected light to computer readable signals. Typically, the spectrometer device further includes means for focusing the excitation light source to a spot and means for positioning the excitation spot on the SERS-active substrate and/or means for spectrally resolved detection of the light scattered from the SERS-active surface. Advantageously, the SERS-active surface is adapted to enhance the Raman signal in so-called hot-spots. The spectrometer device may be configured for localized acquisition of Raman/SERS-spectral data from a surface by any known technique, including advanced techniques such as resonant excitation Raman spectroscopy.

Further according to one embodiment of the system, at least one of the one or more target components is populated during estimation of the parameters of the factorization model by

- a fixed value target spectrum that is kept fixed during estimation/inference;
- a parametric description of the target spectrum with target component parameters that are estimated/inferred by the joint estimation; or
- a probabilistic parametric description of the target spectrum with target component parameters that are estimated/inferred by the joint estimation.

The system may be configured for detecting multiple target components, and more than one of these target components may be subject to respective population rules. The population rules for populating different target components may be of the same type and/or may differ from each other.

By imposing a population rule representing a Raman/SERS-spectral signature of the target analyte on at least some of the target components, forehand information about the target analytes is integrated directly in the detection apparatus, thereby lowering the detection limit and improving the robustness of the detection of that target analyte at very low concentrations. For each target analyte, the forehand information can thus be implemented in a convenient, flexible, and intuitively straightforward manner. Further advantages of the system are evident from the above discussion of the method of detecting one or more target analytes in a sample.

Advantageously, a system according to any of the above-mentioned embodiments is configured to perform the method according to any of the above-mentioned embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

Preferred embodiments of the invention will be described in more detail in connection with the appended drawings, which show in

FIG. 1 a schematic diagram of a system for detecting a target analyte according to one embodiment,

FIG. 2 a SEM micrograph of a SERS-active substrate,

FIG. 3 (a) a Raman/SERS-map and (b) a Raman/SERS-spectrum at a given location on the map,

FIG. 4 (a) a Raman/SERS-map and (b) three spectra at three locations (I, II, III) on the map,

FIG. 5 (a) a first loadings-map and (b) the corresponding first spectral component as estimated in a factorization model of an ensemble X of Raman/SERS-spectra,

FIG. 6 (a) a second loadings-map and (b) the corresponding second spectral component as estimated in the factorization model of Fig. 5, and in

FIG. 7 (A) a Raman/SERS-map, (B) a loadings map, and (C) the corresponding target component as estimated in a factorization model of an ensemble of Raman/SERS-spectra for low concentrations of target analyte.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

FIG.1 shows a schematic diagram of a system 100 for detecting a target analyte in a sample according to one embodiment of the invention. The system comprises a measurement portion 101 and a data analysis portion 102. The measurement portion 101 comprises a SERS-device 110 with a SERS-active surface 111 on which analytes 99 from a sample are captured. The measurement portion 101 further comprises a spectrometer device 120 with an excitation light source 121 adapted to excite Raman scattering events on the SERS-active surface 111, and a detection unit 122 for the spatially and spectrally resolved detection of light scattered from the SERS-active surface 111 and for conversion of the detected light to computer readable signals. Typically, the spectrometer device 120 further includes means for focusing the excitation light from source 121 to a spot and means for positioning the excitation spot on the SERS-active surface 111. Means for positioning the excitation spot for a performing a measurement may include optical means for moving the excitation light with respect to the SERS-surface, means for mechanically positioning the SERS-device 11 and/or mechanically positioning the spectrometer device 120. The spectrometer device 120 is thus configured for the acquisition of an ensemble X of localized/spatially resolved Raman/SERS-spectra from the SERS-active surface 111. The spectrometer device 120 may be configured for acquisition of Raman/SERS-spectral data from a surface by any known Raman-spectroscopy technique, including advanced techniques such as resonant excitation Raman spectroscopy. In the examples given further below, the measurement portion 101 has a SERS-device 110 with an arrangement of silver coated silicon nano-pillars as the SERS-active surface 111 on a silicon substrate. Preparation of such SERS-devices 110 is disclosed e.g. in WO 2011/047690 A1. An SEM-micrograph of silver coated silicon nano-pillars on a SERS-active surface is depicted in Fig.2. The spectrometer device 120 used for acquiring the Raman/SERS-data of the examples is a Thermo Scientific DXR Raman Microscope using a 780 nm excitation laser. A laser powered at 0.5 mW power was used in conjunction with a 50x optical objective which yields

an approximately 1 µm diameter laser spot size. The total acquisition time in each spot was 1 s, wherein each spot was measured twice within that total acquisition time, before a 6^{th} order polynomial baseline correction was performed. Each ensemble of Raman/SERS-spectra consisted of spectra gathered in larger squares (up to 10000 points) with a 1x1 µm grid. The target analyte was 17β-Estradiol (E2) in a fluorescent labelled version called Estradiol Glow (EG), which is available from Jena Bioscience GmbH, Germany. High concentration solutions comprise 1 µM EG and low concentrations comprise 10 nM EG.

The ensemble X of Raman/SERS-spectra is passed to the data analysis portion 102 and stored in a data storage 130 in the form of a matrix [X] of N spectra with D values each. The ensemble X of Raman/SERS-spectra is then passed to a data modelling device 140 for modelling by a factorization model with a loadings matrix A times a spectral component matrix S and an additive term E representing the measurement noise. The spectral component matrix S has K spectral component vectors with D values each, and the loadings matrix A has N loading vectors with K elements for weighting of the K spectral components in a linear combination modelling the total Raman/SERS-signal in each of the N locations/instances. The number K of spectral components is a model parameter that may be chosen beforehand. In the examples given below, K was chosen to be ten. The parameters of the factorization model are estimated in a Bayesian framework using a Gibbs sampler, and the result of the estimation is passed from the data modelling device 140 to a classifier 150 for determining the presence of the target analyte from the parameters of the model.

The spectral component representing the Raman/SERS-spectral signature of the target analyte is referred to as the target component. The target component is selected amongst the spectral components of the spectral component matrix S. The presence of the target analyte may advantageously be established from the loadings of the spectral component representing the Raman/SERS-spectral signature of the target analyte as discussed above. The target component for a given analyte may be selected amongst the group of spectral components $\{S_{free}\}$ with values that may be freely estimated. In this case, the target component is identified by comparison with a known spectral response of the target analyte, such as derived from a spectral data library, from a theoretical model, or experimentally from a measurement on

a reference sample. The comparison with the known spectral response is performed after estimation of the model parameters.

Alternatively, the known spectral response of the target analyte may be integrated directly in the model as forehand information, thereby significantly improving the performance of the data analysis portion in terms of lower detection limit and enhanced robustness of the detection as compared to a model neglecting such forehand information. This is possible, since the target analyte to be tested for is known beforehand. Advantageously, the target component of a given target analyte may be selected in a group of spectral components that are subject to a population rule for populating the values of the spectral components during the estimation routine. For example, the population rule for the target component of the given analyte may be that all values are fixed to a specific spectrum throughout the estimation, thus only allowing the loadings of this specific spectrum to be estimated in the model. The fixed spectrum may be a known spectral response as discussed above. However, due to the strict nature of the fixed spectrum population rule, it is recommended to use a forehand knowledge of the target component that is close to the spectral response of the target analyte under the actual experimental conditions of the measurement portion 101.

In some embodiments, the population rule may describe the target component in a parametric form. In the case of multiple target analytes each of the target components may be subject to a different population rule.

Multiple spectral components subject to such population rules may be grouped as a group of fixed spectral components $\{S_{fix}\}$ and a group of parametric spectral components $\{S_{para}\}$, respectively. The parametric functions describing the parametric spectra may be probabilistic parametric functions. The parameters of the parametric spectral components $\{S_{para}\}$ are estimated along with the remaining parameters of the factorization model. In general, the spectral component matrix S may thus be written as a combination of a “free” portion {Sfree} of spectral components that are not subject to a population rule and optional portions {Sfix}, {Spara} of spectral components that are subject to a population rule: $S = (\{S_{free}\}, \{S_{fix}\}, \{S_{para}\})$.

EXAMPLES

Example 1

FIG.3 shows (a) a Raman/SERS-map and (b) a Raman/SERS-spectrum at a given location on the map as indicated by the arrow. The Raman/SERS-map visualizes the recorded intensities at a specific Raman shift chosen to be where the spectral response of the target molecule is expected to have a peak. Areas with high intensities are identified and these areas are denoted as potential hot-spots containing target molecules. The Raman/SERS data shown in Fig.3 illustrates an uncontaminated Raman/SERS measurement. Fig.3(a) maps the Raman/SERS-signal intensities at 1166 cm^{-1} and Fig.3(b) gives the Raman/SERS-spectrum for Estradiol Glow (EG) at the location where the Raman/SERS-map for 1166 cm^{-1} has a maximum. Hot-spots containing molecules are readily identified on the Raman/SERS-map as the bright areas. In the Raman/SERS-spectrum for EG, the peaks at 1166 cm^{-1} and 1580 cm^{-1} are considered the major discriminative. The spectra are collected over a 30 µm × 30 µm area with a resolution of 1 µm. Traditional analysis of Raman/SERS maps rely on choosing one or two Raman shifts where the dominant peaks for the molecule of interest are present. In the case where the target molecule is present (and at a hot-spot) the intensities are considered to be heavy tail distributed whereas Raman/SERS maps for blank substrates are considered to be normally distributed. Detection is then performed by testing if a Raman/SERS-map has heavy tailed intensities. This approach is not specific and will be sensitive to outliers or contaminants that give high intensities at the selected Raman shifts. This becomes clear when considering FIG.4.

Example 2

FIG.4 gives an example of a Raman/SERS-map that has been locally contaminated. Analogue to the above example of an uncontaminated set of Raman/SERS-data, Fig.4(a) maps the Raman/SERS-signal intensities at 1166 cm^{-1}. The Raman/SERS-map at 1166 cm^{-1} exhibits a number of maxima. Fig.4(b) gives three spectra labelled (I, II, III) corresponding to the three locations of maximum intensity indicated by (I, II, III) on the map. Only the very weak spectrum labelled (III) is actually the Raman/SERS-spectrum for the target analyte EG. The other spectra (I, II) are to be attributed to contaminants, most probably salt residuals stemming from sample preparation steps. The contaminant signals obscure the target Raman/SERS signal,

mislead the traditional data analysis methods and thereby prevent an appropriate detection of the target analyte. As illustrated in Fig.5 and Fig.6, this issue is addressed by using the factorization model according to the present for de-mixing the different spectral components.

Example 3

Fig.5 and Fig.6 illustrate the noise filtering achieved by using the factorization model X = AS + E according to the present invention. A, S and E are jointly estimated in a Bayesian framework using exponential priors with a rate equal to one for both A and S in order to account for the sparse nature of these matrices in analogy to the model described in Schmidt et al., "Bayesian Non-negative Matrix Factorization" (2009), Lecture notes in Computer Science, pp.540-547. In this paper an effective Gibbs sampler was derived that is also used here for parameter estimation. In our experiments, a burn in period of 20 Gibbs-sweeps was used and then 50 sweeps were used to generate 50 samples from each component. The mean values from these samples were used as the parameter estimates. As a prior for the noise variance in E we chose a inverse gamma density with shape and scale parameters equal to zero (a flat improper prior), in order
to let the data dictate the inference regarding the measurement noise. Finally, the number of basis vectors of the spectral component matrix was set to K = 10, to allow for sufficient flexibility to identify rare but signal-intense outliers. Fig.5(a) and Fig.6(a) map first and second loadings for first and second spectral components, respectively. The loadings are extracted from the loadings matrix A. Fig.5(b) and Fig.6(b) show the corresponding first and second spectral component for these first and second loadings as extracted from the spectral component matrix S.

Using the spectrum in S that is identified as an EG spectrum, the most dominant area for 1166cm^{-1} (area I in Fig. 4(a)) has been correctly removed from the map. In addition, the mixed area (spectrum II on Fig.4(b)) has been de-mixed into two distinct components, the EG spectrum and the contaminants. The EG-spectrum is selected as the target component in the spectral component matrix S. The loadings for the target component can now be used to identify where EG is likely to be present in the map. The method correctly identifies spots where EG is present. At high concentrations the proposed method successfully identifies a Raman/SERS-spectrum as

one of the spectral components in S. The method is applied on the data shown in Fig.3, and one of the spectral component basis vectors readily corresponds to a spectrum for EG. The Raman/SERS-spectrum for EG as derived from the freely estimated spectral component matrix S for high concentration measurements may now be used as a fixed target component S_{fix} in a new test for detecting the presence of EG as the target analyte at very low concentrations. As illustrated in Fig.7, the sensitivity and selectivity of the method/system for detecting the target analyte, here EG, is significantly improved as compared to traditional data analysis.

Example 4

For very low concentrations the Raman/SERS-spectrum of interest will only make up a minor portion in the data and the aforementioned approach is often not able to identify the target spectrum. Despite its improved noise filtering properties as compared to the traditional data analysis, at low concentrations, the factorization model is not able to automatically identify the EG spectrum as one of the spectral components in S. This is most probably, because the target molecule is giving a weak signal as well as being a rare occurrence.

This is solved by having the spectra of interest (target components) as fixed vectors in S during parameter estimation. The spectra of interest can for example be learned from high concentration measurements. The remaining components in S are sampled as normal. Fig.7(A) maps the Raman/SERS signal intensity at 1166 cm^{-1}; Fig.7(B) maps the loadings of the EG target component; and Fig.7(C) provides the Raman/SERS-spectrum at a location that is identified by the data-analysis as having EG present. Using the spectral component identified as EG in the high concentration measurements shown on Fig.5 as a vector in S when analysing the ensemble of Raman/SERS-spectra from the low concentration sample, and keeping that column fixed during parameter estimation (probing mode), areas with EG on the SERS-active substrate are successfully identified.

CLAIMS

1. Method of detecting the presence of one or more specific target analytes in a sample, the method comprising the steps of
   (a) capturing analytes from the sample on a SERS-active surface;
   (b) acquiring an ensemble X of Raman/SERS-spectra measured on the SERS-active surface;
   (c) for each target analyte, determining a target spectrum representative of that target analyte;
   (d) modelling the ensemble X of measured Raman/SERS-spectra using a factorization-model, wherein the factorization-model includes a loadings matrix A times a spectral component matrix S, wherein matrix elements of A and S are all real and non-negative;
   (e) jointly estimating parameters in the factorization-model, the estimated parameters including at least A and S;
   (f) for each target analyte, selecting amongst the spectral components of S a target component corresponding to the target spectrum of that target analyte; and
   (g) determining presence of the one or more target analytes at least on the basis of the estimated loadings for the selected target components;
   wherein estimation of at least one of the one or more target components is subject to a population rule determining the population of the target component elements with values according to a pre-determined function.

2. Method according to claim 1, wherein the predetermined function is one of
   - a fixed value target spectrum that is kept fixed during estimation/inference;
   - a parametric description of the target spectrum with target component parameters that are estimated/inferred by the joint estimation; and
   - a probabilistic parametric description of the target spectrum with target component parameters that are estimated/inferred by the joint estimation;

3. Method according to claim 1 or claim 2, wherein at least a first target component and a second target component are subject to respective population rules,

wherein the predetermined functions determining the population of the first and second target components are different from each other.

4. Method according to any of the preceding claims, wherein the target spectrum of at least one target analyte in step (c) is determined from Raman/SERS measurements of said at least one target analyte captured on a SERS-active surface.

5. Method according to claim 4, wherein the target analyte is captured on the SERS-active surface by exposing the SERS-active surface to a reference sample comprising the target analyte.

6. Method according to claim 5, wherein the Raman/SERS measurements include an ensemble R of Raman/SERS-spectra, and the target spectrum is extracted from a factorization-model of the ensemble R.

7. Method according to any of the preceding claims, wherein the factorization-model further includes an additive term with a residual noise matrix E, wherein matrix elements of E are all real.

8. Method according to any of the preceding claims, wherein the factorization-model further includes an additive term with a background matrix B, wherein matrix elements of B are all real and non-negative.

9. Method according to any of the preceding claims, wherein the joint estimation is performed in a Bayesian framework.

10. Method according to any of the preceding claims, wherein a total number K of spectral components forming the spectral component matrix S is treated as a parameter of the factorization model, which is estimated by the joint estimation, preferably using non-parametric Bayesian techniques.

11. Method according to any of the preceding claims, wherein a detection result for the presence of a given target analyte is determined as positive if the loading of the target component exceeds a pre-determined threshold.

12. Method according to any of the preceding claims, wherein a positive detection result for the presence of a given target analyte in one location on the SERS-active surface is validated by regarding loadings of the corresponding target component in neighbouring locations on the SERS-active surface.

13. System for detecting the presence of one or more specific target analytes in a sample, the system comprising a SERS-device with a SERS-active surface, a spectrometer device configured for the spatially and spectrally resolved acquisition of Raman/SERS-spectral data from the SERS-active surface, and a computer including devices for data storage and data processing, wherein the computer is configured for performing the steps of

(a) storing an ensemble X of measured Raman/SERS-spectra;

(b) modelling the ensemble X of measured Raman/SERS-spectra using a factorization-model, wherein the factorization-model includes a loadings matrix A times a spectral component matrix S, wherein the matrix elements of A and S are all real and non-negative;

(c) jointly estimating parameters in the factorization-model, the estimated parameters including at least A and S

(d) for each target analyte, selecting amongst the spectral components of S a target component corresponding to a pre-determined target spectrum for that target analyte; and

(e) for each target analyte, determining presence of the target analyte at least on the basis of the estimated loadings for the corresponding target component;

wherein during joint estimation of the parameters of the factorization model at least one of the one or more target components is populated according to a predetermined function.

14. System according to claim 13, wherein the predetermined function is

- a fixed value target spectrum that is kept fixed during estimation/inference;
- a parametric description of the target spectrum with target component parameters that are estimated/inferred by the joint estimation; or

- a probabilistic parametric description of the target spectrum with target component parameters that are estimated/inferred by the joint estimation.

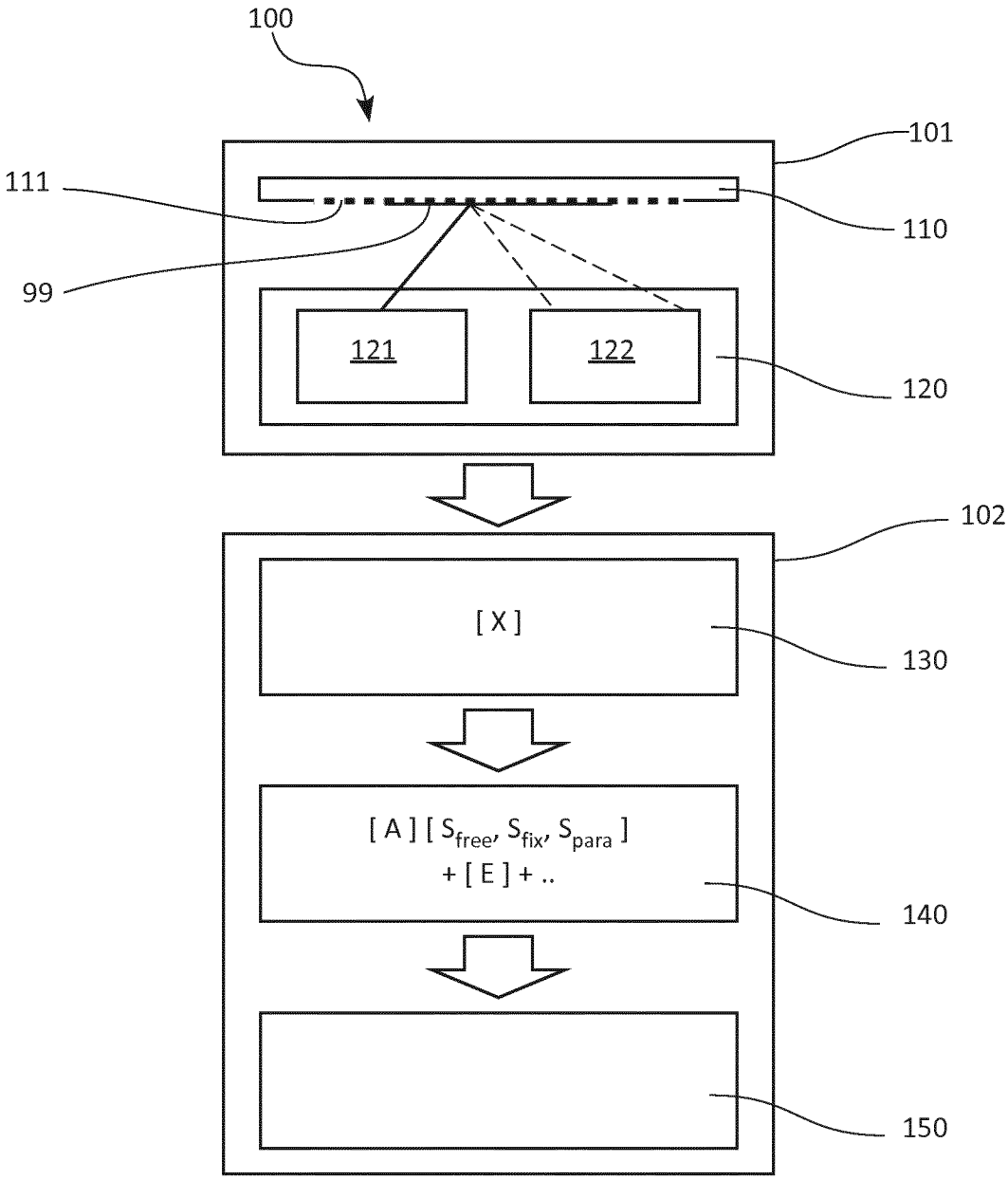
100
101
111
110
99
121
122
120
102
[X]
130
[A] [S_{free}, S_{fix}, S_{para}]
+ [E] + ..
140
150

Fig. 1

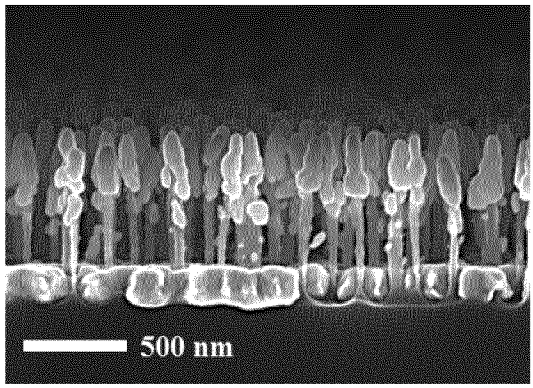

Fig. 2

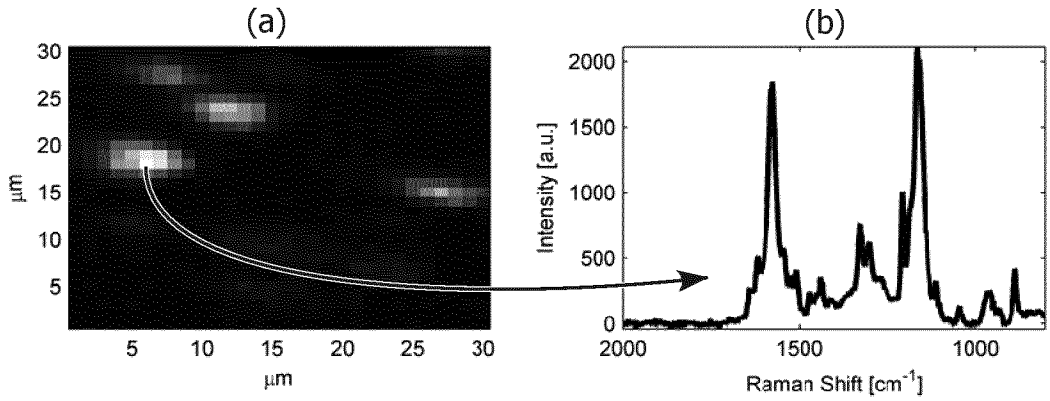

Fig. 3

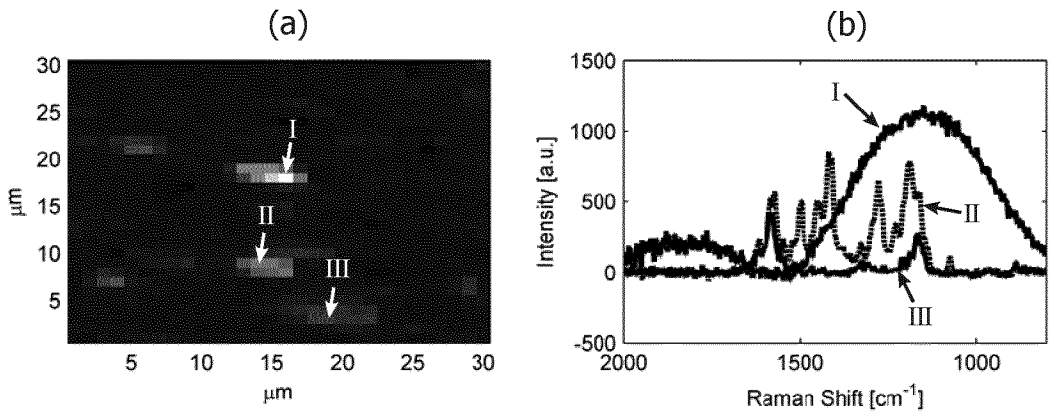

Fig. 4

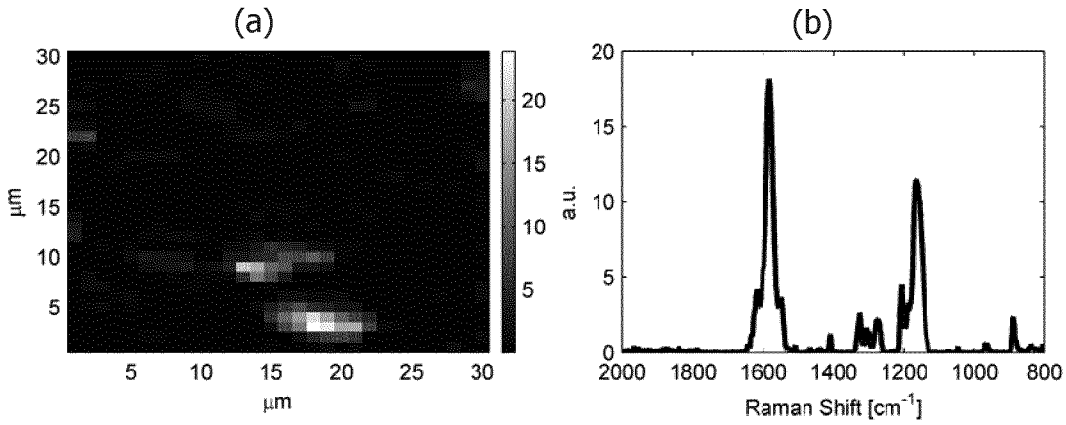

Fig. 5

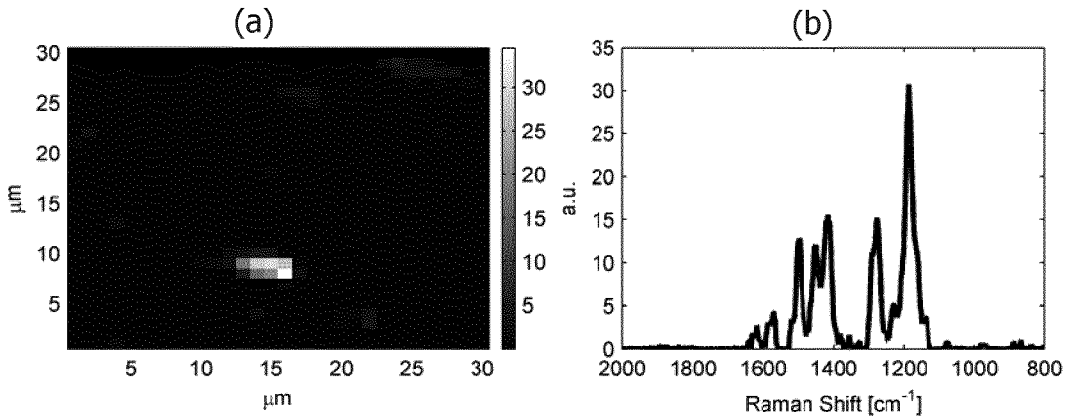

Fig. 6

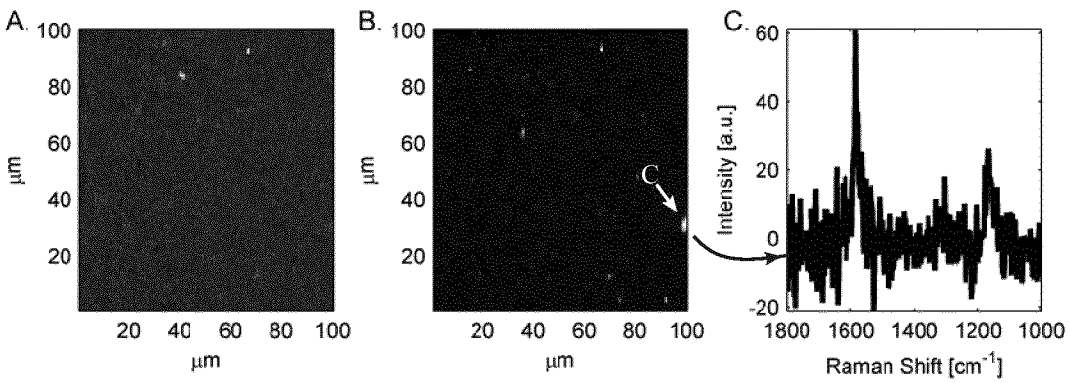

Fig. 7

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2015/070723

A. CLASSIFICATION OF SUBJECT MATTER
INV. G01N21/65
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, COMPENDEX, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	TAMÁS FIRKALA ET AL: "Investigation of drug distribution in tablets using surface enhanced Raman chemical imaging", JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS, vol. 76, 1 March 2013 (2013-03-01), pages 145-151, XP055225931, US ISSN: 0731-7085, DOI: 10.1016/j.jpba.2012.12.017 paragraphs [0002], [0003] figures 4,5 ----- -/--	1-14

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
9 November 2015	16/11/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Rasmusson, Marcus

2

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MÓNICA B. MAMIÁN-LÓPEZ ET AL: "Quantification of moxifloxacin in urine using surface-enhanced Raman spectroscopy (SERS) and multivariate curve resolution on a nanostructured gold surface", ANALYTICAL AND BIOANALYTICAL CHEMISTRY, vol. 405, no. 24, 1 September 2013 (2013-09-01), pages 7671-7677, XP055225853, DE ISSN: 1618-2642, DOI: 10.1007/s00216-013-7200-y the whole document	1-14
A	ZHOU KENNETH J ET AL: "Nonnegative matrix factorization: a blind sources separation method to extract content of fluorophores mixture media", PROGRESS IN BIOMEDICAL OPTICS AND IMAGING, SPIE - INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING, BELLINGHAM, WA, US, vol. 8947, 4 March 2014 (2014-03-04), pages 89471V-89471V, XP060033874, ISSN: 1605-7422, DOI: 10.1117/12.2035536 ISBN: 978-0-8194-9850-2 abstract	1-14