UNIVERSITY OF LIMERICK OLLSCOIL LUIMNIGH | Bernal Institute

www.bernalinstitute.com

Analysis of Raman Spectra Signals Based on Molecular Fingerprints from DFT data

Milad Asgarpour Khansary, Saeed Shirazian, Gavin Walker

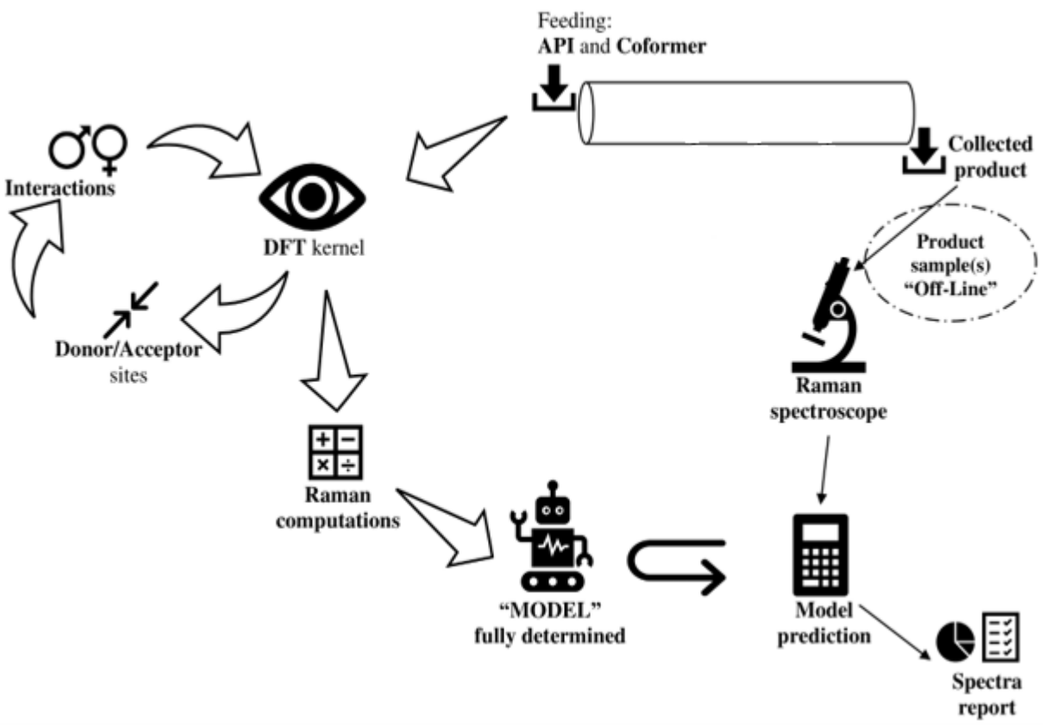
Abstract

In cocrystalization-based processing of pharmaceuticals, Raman spectra is utilized as a standard tool for quality control. However, such analysis suffers from the lack of relevant tools for realtime analysis of spectra. Here we report an automation method to facilitate such analysis. For this purpose, we coupled the nonlinear multivariable functional series approximation with molecular fingerprints as derived from density functional theory calculations. As the case study, we considered the cocrystalization process of ibuprofen and nicotinamide. We showed that for any mixture, it is possible to identify the contribution (probability) of different molecular fingerprints by just reading the Raman spectra, with significant level of confidence. Therefore, a realtime monitoring of continuous cocrystalization-based pharmaceuticals processing is achieved.

Data availability: The associated data of this paper can be found under Downloads page at: https://sites.google.com/view/makhansary, using tag ID = CoCryM.ExSysR.

Computations

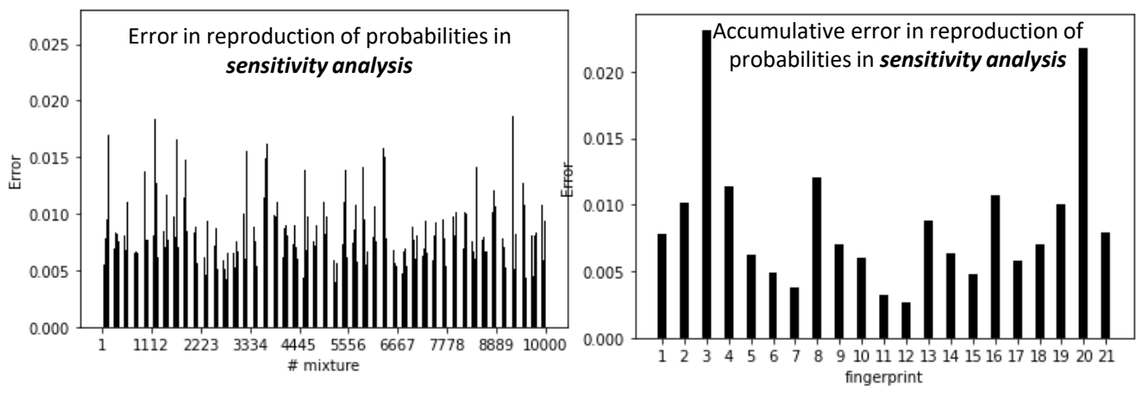
The matrix of fingerprints, ρ needs Raman intensities as determined from density functional theory calculations. Raman spectra can be calculated for any incident light and system temperature using the data from density functional theory following the method of Porezag and Pederson. We performed relaxations employing the Generalized Gradient Approximations with Perdew-Becke-Ernzerhof functional including implicit solvent as described by COnductor-like Screening MOdel. To control the convergence behaviour for an enhanced self-consistent field calculation, thermal smearing was also applied including double numerical basis including d-polarization function level of theory. The convergence tolerances are energy: 2.0×10^{-5} kcal/mol, force: 10^{-3} kcal/mol/Å, max iterations: 10^4, displacement: 10^{-5} Å. The matrix ρcan be singular due to nature of Raman intensities, so we compute its Moore-Penrose pseudo-inverse using a least-squares solver (*linalg.pinv* in NumPy package). ***Case study***: We considered the cocrystalization process of ibuprofen and nicotinamide by Raman spectra from literature, where they carried out measurements with 532 nm excitation at a laser power of 150mW, over 200 to 1800 cm^{-1}. ***Sensitivity analysis***: we generated 10^4 random dummy mixtures (R_d) as $R_d = \sum_{i=A,B,1}^{N} rnd_i \times r_i$ where rnd_i is random number acting as the weight/probability. The difference between calculated linear weights of fingerprints, $a_i'$from model and rnd_i is defined as the model error. All Raman intensities are normalized as $\overline{R} = \frac{R-min(R)}{max(R)-min(R)}$.

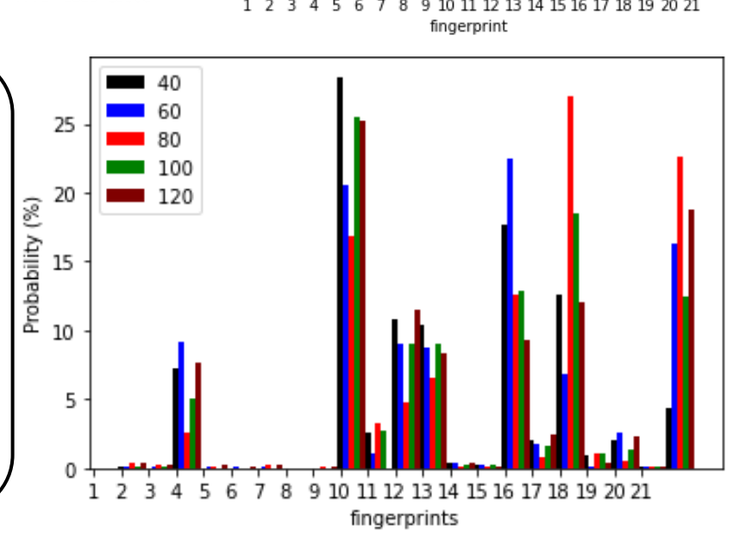


The workflow of computations in model development and its application

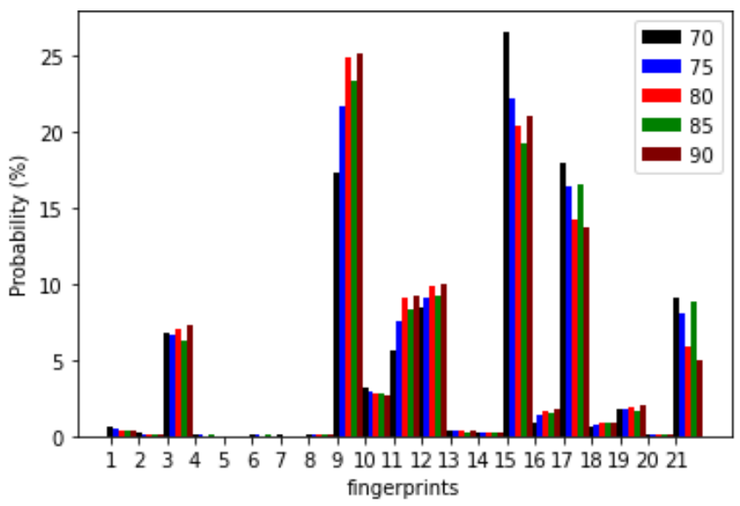
Model details

We assume that a nonlinear (weak) correlation should exist among measured Raman spectra as read by spectrometer and Raman spectra of fingerprints (pairs of molecule *A* and *B*), no matter identification of which fingerprint is target of measurements. We show Raman spectra of target of measurement by R and the Raman spectra of molecule *A,* molecule *B*, and all fingerprints by r_i, considering a total of *N* fingerprints (different donor/acceptor interactions). The nonlinear correlation, f, of measured Raman intensity R to fingerprint Raman intensities, r_i can be given as $R = f(r_A, r_B, r_1, r_2, \dots, r_N)$. Here, the exact form of nonlinear correlation, f, is not accessible and of no concern. But it is possible to approximate$f$, by a new function $\overline{f}$to represent the relationship between the Raman intensity $\overline{R}$ and Raman intensity of fingerprints, r_i as $\overline{R} = \overline{f}(r_A, r_B, r_1, r_2, \dots, r_N)$. Employing volterra functional series approximation approach, we show that this correlation reads $\overline{R} = a + \sum_{i=A,B,1}^{N} a_i r_i + \sum_{i=A,B,1}^{N}\sum_{j=A,B,1}^{N} a_{ij} r_i r_j + \sum_{i=A,B,1}^{N}\sum_{j=A,B,1}^{N}\sum_{k=A,B,1}^{N} a_{ijk} r_i r_j r_k$ where a, a_i , a_{ij} , a_{ijk} are all unknow constant coefficients. We write this equation in matrix form as $R = \rho A$ which can be solved as $A = (\rho^T \rho)^{-1} \rho^T R$. ρ is the matrix containing all the fingerprints r_i. R is the vector of Raman intensity of target, $A$is a vector of all coefficients a, a_i , a_{ij} , a_{ijk}. Using coefficients a_i , we define intuitive linear weights of each fingerprint a_i' in signal once as $a_i' = \frac{a_i}{[\sum_{i=A,B,1}^{N} a_i] - a_A - a_B} \quad for \quad i = 1 \to N$.





Prediction performance of model in estimating probability of fingerprints in Raman data for mixture extruded at screw speeds 40, 60, 80, 100, and 120 rpm at 90°C



Prediction performance of model in estimating probability of fingerprints in Raman data or mixture extruded at temperatures 70, 75, 80, 85, and 90 °C screwed at 40 rpm

Notes

Int J Pharm, 2020. 591: p. 119992.

IEEE Transactions on Systems, Man, and Cybernetics, 1971. 1(4): p. 364-378.

Physical Review B, 1996. 54(11): p. 7830-7836.

Physical Review Letters, 1996. 77(18): p. 3865-3868.

Int J Pharm, 2019. 566: p. 745-755.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 801165 with co-funding through Confirm smart manufacturing centre by its funding body, Science Foundation Ireland (SFI), grant no. 16/rc/3918.

Marie Skłodowska-Curie Actions | Science Foundation Ireland sfi For what's next | Confirm Smart Manufacturing | PROCESS Engineering Solutions for the Process Industry