Supporting Information

Dynamic Sparse Sampling for Confocal Raman Microscopy

Shijie Zhang[a], Zhengtian Song[a], G. M. Dilshan P. Godaliyadda[b], Dong Hye Ye[b], Azhad U. Chowdhury[a], Atanu Sengupta[c], Gregery T. Buzzard[d], Charles A. Bouman[b] and Garth J. Simpson[a]*

[a]Department of Chemistry, Purdue University, West Lafayette, IN, 47907.
[b]Department of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, 47097.
[c]Dr. Reddy's Laboratories, IPDO, Bachupally Campus, Hyderabad, Telengana-500090, India
[d]Department of Mathematics, Purdue University, West Lafayette, IN, 47097.

*Correspondence should be addressed to: Prof. Garth J. Simpson (gsimpson@purdue.edu)

Captions for Supporting Videos:

Video S1. Demonstration movie of simulated SLADS Raman imaging process, based on ground truth spectra

Video S2. Demonstration movie of experimental SLADS Raman measurement process

This section provides details of the method used to prepare training images of supervised learning approach for dynamic sampling (SLADS) for stopping conditions. Training data were prepared by processing bright field (BF) images and second harmonic generation (SHG) images in combination with BF images acquired with the dynamic sampling Raman instrument. As demonstrated in previous work {Chowdhury, 2017 #7}. SHG provides an independent method of determining particle polymorphism with high confidence.

Training images integrating BF and SHG are shown in **Figure S1**. First, individual particles of clopidogrel bisulfate polymorphs Form I (bright in SHG image) and Form II (relatively dim in SHG image) were manually identified and marked. Description of the sample preparation can be found in the body of the manuscript. Correlation of the fields of view (FOVs) for both BF and SHG was accomplished by cropping and resizing the BF images. Next, the resized BF images were converted into binary images by thresholding to differentiate clopidogrel bisulfate particles from the background. Then, using the knowledge obtained from the first step, within these binary images, Form I and Form II particles were manually classified (indicated in red and blue, respectively). Eventually, five discrete-valued images with a resolution of 128 px × 128 px were finalized and used as training data for the dynamic sampling Raman, shown in **Figure S2**.

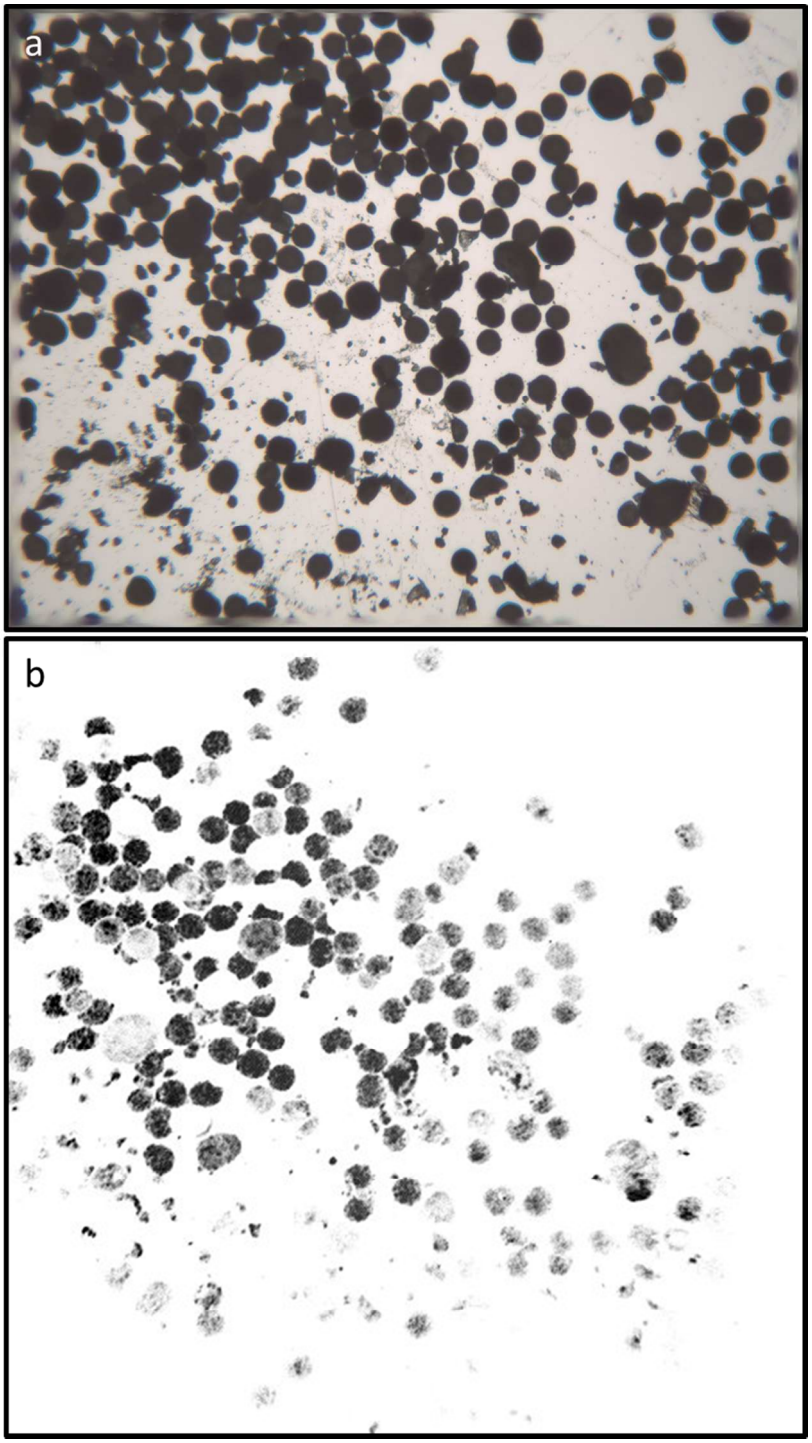

Figure S1. (a) BF image and (b) SHG image of a clopidogrel bisulfate sample acquired using SHG. Image borders were added to mark out the FOVs of the two imaging modalities. In the BF image, both Form I and Form II polymorphs are shown as dark particles with a spherical shape. In the SHG image, Form I polymorph is shown as dark particles, and Form II polymorph does not give signals.

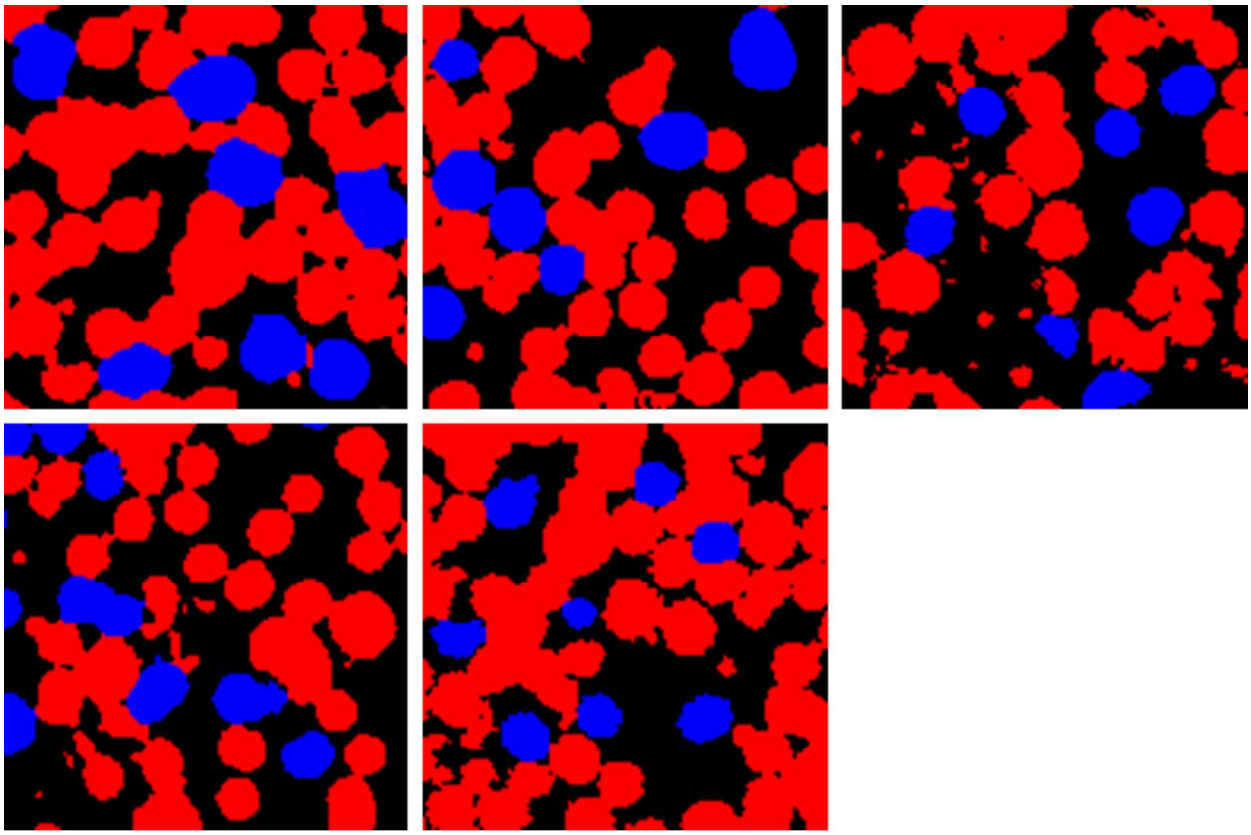

Figure S2. The five discrete valued images cropped and processed from **Figure S1(a)**, and used as training images for dynamic sampling Raman imaging stopping conditions.