Complementary classification of solid oral dosage forms in ambient conditions by desorption electrospray ionization mass spectrometry and transmission Raman spectroscopy

This manuscript ([permalink](#)) was automatically generated from [adamjtaylor/tablet-classification@72dcec7](#) on October 14, 2021.

Authors

- **Adam J Taylor**
  ⓘ [XXXX-XXXX-XXXX-XXXX](#) · [adamjtaylor](#) · [adamjtaylor](#)
  National Physical Laboratory, Teddington, UK; Sage Bionetworks, Seattle, WA, USA

- **Dimitrios Tsikritsis**

  National Physical Laboratory, Teddington, UK

- **Alex Dexter**

  National Physical Laboratory, Teddington, UK

- **Amy Burton**

  National Physical Laboratory, Teddington, UK

- **Josephine Bunch**

  National Physical Laboratory, Teddington, UK; Imperial College London, London, UK; The Rosalind Frankilin Insitute, Harwell, UK

- **Natalie Belsey**

  National Physical Laboratory, Teddington, UK

Abstract

Discrepancies or defects in active ingredients, excipients and coatings that form solid oral dosage forms can both impact product quality and provide hallmarks of off-brand or counterfeit products. There is therefore a need for rapid and continuous analytical techniques that can assess and classify product differences of intact samples at- or near the production line, or in analytical labs, ideally without resorting to product dissolution.

Here we test the ability of two rapid ambient chemical characterization methods to discriminate between solid dosage forms: desorption electrospray ionization mass spectrometry and transmission Raman spectroscopy. These two techniques are highly complementary, offering greater sensitivity to the analysis of the surface and the tablet bulk, respectively. The data sets generated were then used to test a variety of classification algorithms including linear discriminate analysis, tree-based methods, a simple neural network, and support vector machines (SVM). The highest performing algorithms for DESI-MSI were the SVM, with an additional performance boost when used with a polynomial kernel. For transmission Raman data, a linear discriminant analysis (LDA) model was found to be the most effective.

Introduction

Inconsistencies in active ingredients, excipients, the thickness and integrity of coatings and the presence of impurities in solid oral dosage forms all negatively affect their performance. Inferior quality attributes can be useful to identify off-brand or counterfeit products. There is a need for rapid and continuous analytical techniques that can assess and classify product differences of intact samples at- or near the production line, or in analytical labs, ideally without resorting to product dissolution [1]. Rapid measurement tools are particularly important to enable continuous monitoring, necessary to support the change from batch to continuous manufacturing. Analytical methods are required to monitor both the actives, coatings and consistency of the product: For example, in addition to the total API content, insight is also needed on degradation products, impurities, (co-) crystallinity/presence of polymorphs, and content uniformity. The ability to monitor tablet coating thickness and integrity is of great importance, particularly for functional coatings, such gastro-resistance, which would be compromised by insufficient thickness, or the occurrence of cracks in the film [2].

This manuscript is a template (aka "rootstock") for Manubot, a tool for writing scholarly manuscripts. Use this template as a starting point for your manuscript.

The rest of this document is a full list of formatting elements/features supported by Manubot. Compare the input (`.md` files in the `/content` directory) to the output you see below.

Basic formatting

Bold text

Semi-bold text

Centered text

Right-aligned text

Italic text

Combined *italics and **bold***

~~Strikethrough~~

1. Ordered list item
2. Ordered list item
   a. Sub-item
   b. Sub-item
      i. Sub-sub-item
3. Ordered list item
   a. Sub-item

- List item
- List item
- List item

subscript: H_2O is a liquid

superscript: 2^{10} is 1024.

[unicode superscripts]^{0123456789}

[unicode subscripts]$_{0123456789}$

A long paragraph of text. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

Putting each sentence on its own line has numerous benefits with regard to [editing] and [version control].

Line break without starting a new paragraph by putting
two spaces at end of line.

Document organization

Document section headings:

Heading 1

Heading 2

Heading 3

Heading 4

Heading 5

Heading 6

A heading centered on its own printed page

Horizontal rule:

`Heading 1`'s are recommended to be reserved for the title of the manuscript.

`Heading 2`'s are recommended for broad sections such as *Abstract*, *Methods*, *Conclusion*, etc.

`Heading 3`'s and `Heading 4`'s are recommended for sub-sections.

Links

Bare URL link: https://manubot.org

Long link with lots of words and stuff and junk and bleep and blah and stuff and other stuff and more stuff yeah

Link with text

Link with hover text

Link by reference

Citations

Citation by DOI [3].

Citation by PubMed Central ID [4].

Citation by PubMed ID [5].

Citation by Wikidata ID [6].

Citation by ISBN [7].

Citation by URL [8].

Citation by alias [9].

Multiple citations can be put inside the same set of brackets [3,7,9]. Manubot plugins provide easier, more convenient visualization of and navigation between citations [4,5,9,10].

Citation tags (i.e. aliases) can be defined in their own paragraphs using Markdown's reference link syntax:

Referencing figures, tables, equations

Figure 1

Figure 2

Quotes and code

> Quoted text

> Quoted block of text
>
> Two roads diverged in a wood, and I—
> I took the one less traveled by,
> And that has made all the difference.

Code `in the middle` of normal text, aka `inline code`.

Code block with Python syntax highlighting:

```python
from manubot.cite.doi import expand_short_doi

def test_expand_short_doi():
    doi = expand_short_doi("10/c3bp")
    # a string too long to fit within page:
    assert doi == "10.25313/2524-2695-2018-3-vliyanie-enhansera-copia-i-
        insulyatora-gypsy-na-sintez-ernk-modifikatsii-hromatina-i-
        svyazyvanie-insulyatornyh-belkov-vtransfetsirovannyh-geneticheskih-
        konstruktsiyah"
```

Code block with no syntax highlighting:

```
Exporting HTML manuscript
Exporting DOCX manuscript
Exporting PDF manuscript
```

Figures

Figure 1: A square image at actual size and with a bottom caption. Loaded from the latest version of image on GitHub.

Figure 2: An image too wide to fit within page at full size. Loaded from a specific (hashed) version of the image on GitHub.

Figure 3: A tall image with a specified height. Loaded from a specific (hashed) version of the image on GitHub.

Figure 4: A vector `.svg` image loaded from GitHub. The parameter `sanitize=true` is necessary to properly load SVGs hosted via GitHub URLs. White background specified to serve as a backdrop for transparent sections of the image.

Tables

Table 1: A table with a top caption and specified relative column widths.

Bowling Scores	Jane	John	Alice	Bob
Game 1	150	187	210	105
Game 2	98	202	197	102
Game 3	123	180	238	134

Table 2: A table too wide to fit within page.

	Digits 1-33	Digits 34-66	Digits 67-99	Ref.
pi	3.14159265358979323 846264338327950	28841971693993751 0582097494459230	78164062862089986 2803482534211706	`piday.org`
e	2.71828182845904523 536028747135266	24977572470936999 5957496696762772	40766303535475945 7138217852516642	`nasa.gov`

Table 3: A table with merged cells using the `attributes` plugin.

	Colors	
Size	**Text Color**	**Background Color**
big	blue	orange
small	black	white

Equations

A LaTeX equation:

$$\int_0^{\infty} e^{-x^2} dx = \frac{\sqrt{\pi}}{2} \tag{1}$$

An equation too long to fit within page:

$$\begin{aligned} x = a + b + c + d + e + f + g + h + i + j + k + l + m + n + o + p + q + r + s + t \\ + u + v + w + x + y + z + 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 \end{aligned} \tag{2}$$

Special

⚠ **WARNING** *The following features are only supported and intended for* `.html` *and* `.pdf` *exports. Journals are not likely to support them, and they may not display correctly when converted to other formats such as* `.docx`*.*

[LINK STYLED AS A BUTTON]

Adding arbitrary HTML attributes to an element using Pandoc's attribute syntax:

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Adding arbitrary HTML attributes to an element with the Manubot `attributes` plugin (more flexible than Pandoc's method in terms of which elements you can add attributes to):

> Manubot Manubot Manubot Manubot Manubot. Manubot Manubot Manubot Manubot. Manubot Manubot Manubot. Manubot Manubot. Manubot.

Available background colors for text, images, code, banners, etc:

white `lightgrey` `grey` `darkgrey` `black` `lightred` `lightyellow` `lightgreen` `lightblue` `lightpurple` `red` `orange` `yellow` `green` `blue` `purple`

Using the Font Awesome icon set:

✔ ? ★ 🔔 ✖ ⋯

🏳 Light Grey Banner
useful for *general information* - manubot.org

ⓘ Blue Banner
useful for *important information* - manubot.org

🚫 Light Red Banner
useful for *warnings* - manubot.org

References

1. **A Review of PAT Strategies in Secondary Solid Oral Dosage Manufacturing of Small Molecules**
Stephan Laske, Amrit Paudel, Otto Scheibelhofer, Stephan Sacher, Theresa Hoermann, Johannes Khinast, Adrian Kelly, Jukka Rantannen, Ossi Korhonen, Fanny Stauffer, … Brenda Colegrove
Journal of Pharmaceutical Sciences (2017-03) https://doi.org/f965kb
DOI: http://dx.doi.org/10.1016/j.xphs.2016.11.011

2. **PAT-tools for process control in pharmaceutical film coating applications**
Klaus Knop, Peter Kleinebudde
International Journal of Pharmaceutics (2013-12) https://doi.org/f5hg6m
DOI: http://dx.doi.org/10.1016/j.ijpharm.2013.01.062

3. **Sci-Hub provides access to nearly all scholarly literature**
Daniel S Himmelstein, Ariel Rodriguez Romero, Jacob G Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, Casey S Greene
eLife (2018-03-01) https://doi.org/ckcj
DOI: 10.7554/elife.32822 · PMID: 29424689 · PMCID: PMC5832410

4. **Reproducibility of computational workflows is automated using continuous analysis**
Brett K Beaulieu-Jones, Casey S Greene
Nature biotechnology (2017-04) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6103790/
DOI: 10.1038/nbt.3780 · PMID: 28288103 · PMCID: PMC6103790

5. **Bitcoin for the biological literature.**
Douglas Heaven
Nature (2019-02) https://www.ncbi.nlm.nih.gov/pubmed/30718888
DOI: 10.1038/d41586-019-00447-9 · PMID: 30718888

6. **Plan S: Accelerating the transition to full and immediate Open Access to scientific publications**
cOAlition S
(2018-09-04) https://www.wikidata.org/wiki/Q56458321

7. **Open access**
Peter Suber
MIT Press (2012)
ISBN: 9780262517638

8. **Open collaborative writing with Manubot**
Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter
Manubot (2020-05-25) https://greenelab.github.io/meta-review/

9. **Opportunities and obstacles for deep learning in biology and medicine**
Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, … Casey S Greene
Journal of The Royal Society Interface (2018-04-04) https://doi.org/gddkhn
DOI: 10.1098/rsif.2017.0387 · PMID: 29618526 · PMCID: PMC5938574

10. **Open collaborative writing with Manubot**

Daniel S Himmelstein, Vincent Rubinetti, David R Slochower, Dongbo Hu, Venkat S Malladi, Casey S Greene, Anthony Gitter