Data-driven Molecular Engineering of Functional Materials

Jacqueline M. Cole,[a,b,c]

[a] Cavendish Laboratory, Department of Physics, University of Cambridge, UK.

[b] Department of Chemical Engineering & Biotechnology, University of Cambridge, UK

[c] ISIS Neutron and Muon Facility, STFC Rutherford Appleton Laboratory, UK.

The world needs new materials to stimulate industry in key sectors of our economy: environment and sustainability, information storage, optoelectronics, efficiency of chemical processes. Yet, nearly all functional materials are still discovered by 'trial-and-error', whose lack of predictability affords a major materials bottleneck to technological innovation. The emerging field of data-driven molecular engineering offers a prospective solution to this problem; thereby, systematic molecular design and engineering strategies are encoded into algorithms that search through massive {chemical-property} datasets to discover a material that suits a bespoke application. Such data-science approaches to materials discovery are only just becoming possible, given recent advances in artificial intelligence, rapid rises in high-performance-computing capacities, and changes in government legislation that regulates the open-access of scientific data. This talk will present a range of data-driven materials-by-design capabilities that are being developed by the Molecular Engineering group at Cambridge, to accelerate the discovery of new materials. The materials discovery pipeline features the coupling of niche database auto-generation tools, such as ChemDataExtractor [1,2], with machine learning capabilities and the custom design of algorithms to predict new functional materials; we then experimentally validate the predictions using a range of advanced materials characterisation [3] and device testing methods. The discovery of new light-harvesting materials for dye-sensitized solar cells will act as a case study to illustrate these design-to-device methods.

References

[1] M. C. Swain, J. M. Cole, ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature, *J. Chem. Inf. Model.,* **2016**, *56* (10), pp 1894–1904.

[2] C. J. Court, J. M. Cole, Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction, *Scientific Data*, **2017**, 5:180111 | DOI: 10.1038/sdata.2018.111

[3] J. McCree-Grey, J. M. Cole, S. A. Holt, P. J. Evans, Y. Gong, Dye···TiO_2 Interfacial Structure of Dye-Sensitised Solar Cell Working Electrodes Buried under a Solution of I^-/I_3^- Redox Electrolyte, *Nanoscale*, **2017**, 9, 11793-11805.