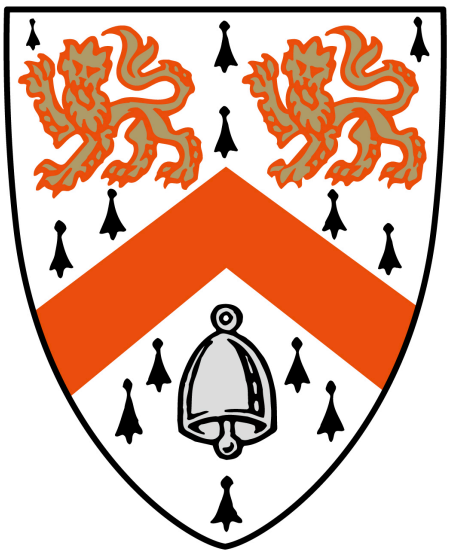
Application of Radial Distribution Functions to Diffraction and Imaging Data: Interfacial Structures, Amorphous, Disordered Materials

Karim Tanju Mukaddem
Wolfson College
University of Cambridge



Department of Physics
This dissertation is submitted for the degree of Doctor of Philosophy.
Date of Submission: December 2019

Declarations

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution.

It does not exceed the prescribed word limit of 60,000 words set by the Department of Physics.

Summary

Application of Radial Distribution Functions to Diffraction and Imaging Data: Interfacial Structures, Amorphous, Disordered Materials

Karim Tanju Mukaddem

The central theme of this thesis is the application of radial and pair distribution function analysis to materials characterisation problems for nanotechnology. These concepts are introduced in Chapter 1, and the associated methods are described in Chapter 2.

Chapter 3 details the first of the results which discusses the design and development of a software tool called *ImageDataExtractor*. This auto-extracts microscopy images and then analyses them to afford quantitative information regarding particles in a sample, such as shape, size and distribution. It realises an opportunity for data-mining the ubiquity of readily available images in the literature. Chapter 4 presents results of the development and execution of a novel experimental technique, called glancing-angle pair distribution function (gaPDF) analysis, applied to the structure of the working electrode in dye-sensitised solar cells (DSSCs). This structure was successfully observed, validating this novel method. The investigation also suggested preferred binding modes of the carboxylic acid anchoring groups present in this interfacial structure. Chapters 5 and 6 demonstrate the application of PDF analysis to synchrotron-based powder diffraction data of two material case studies: the rare earth phosphate glass (REPG) $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$, and four Ru based photo-isomers. The closest R···R rare earth separation, which governs optical properties of REPGs, was determined to be 4.2(1) Å, aided by various statistical techniques. Analysis on four Ru-based photo-isomers confirmed: the existence of local structure in such compounds, their ability to be photo-isomerised in powder form, the theoretical models constructed using computational techniques, and the lack of heterogeneity in photo-isomerisation throughout a given light-induced sample. Chapter 7 concludes the work and offers a future outlook.

Acknowledgements

I would like to thank my supervisor Professor Jacqueline Cole who has been a continuous source of knowledge and support throughout this journey.

Special thanks to my friends in the Molecular Engineering Group, Othman Al Bahri, Edward Beard, Leon Devereux, Apoorv Jain, Jonathan McCree-Grey, and Batuhan Yildirim for all their help.

The Research Complex at Harwell at the Rutherford Appleton Laboratory, UK, is also acknowledged for access to facilities and staff support, with funding being provided by STFC.

Thanks to Dr. Sam Callear and Dr. Alex Hannon of the STFC ISIS Disordered Materials Laboratory at the Rutherford Appleton Laboratory, UK, and Dr. Philip Chater from the Diamond Light Source, UK.

Above all I would like to thank my family, whose emotional and mental support always kept me moving forward through difficult times.

Publications

Chapters 3 and 5 comprise research that has already been published:

Mukaddem, K.T., Beard, E.J., Yildirim, B., Cole, J.M., 2019. ImageDataExtractor: A Tool to Extract and Quantify Data from Microscopy Images. *Journal of Chemical Information and Modeling*. https://doi.org/10.1021/acs.jcim.9b00734, Articles ASAP (Published online: November 12, 2019)

Cole, J.M., Cramer, A.J., Shastri, S.D., Mukaddem, K.T., Newport, R.J., 2018. Multiple rare-earth ion environments in amorphous $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$ revealed by gadolinium K-edge anomalous x-ray scattering. *Physical Review Materials*, *2*(4), p.045604.

Table of Contents

1. Introduction

One of the most challenging stages in the development of new materials is their structural characterisation. The unequivocal determination of the structures of such materials is crucial for subsequent research and development, through the establishment of structure-property relationships.

Certain materials are sufficiently simple and homogeneous that their average surface or bulk structure can be considered representative and used for the establishment of these structure-property relationships. Other materials, usually those that are more specific in their functionality, require the elucidation of local structural information that pertains to the nanometre-to-angstrom scale in order to uncover the mechanisms that underlay their mode of operation. The performance of such materials can be affected by, and sometimes rely on, factors pertaining to their local structure, such as nanoscale morphology, surface reconstructions, interfaces, point or extended defects and others.[1]

An example of the former class of materials would be a crystalline sample of nickel, whose structure can be characterised with sufficient accuracy using bulk analysis techniques (e.g. single-crystal X-ray diffraction), as its functionality is indifferent to local structural perturbations.

An example of the latter class of materials would be trans-$[Ru(NH_3)_4(SO_2)Cl]Cl$ (one of the materials investigated in Chapter 7). The mechanism of functionality exhibited by this material depends on light-induced structural changes at the angstrom scale, and thus it cannot be characterised by a spatiotemporal average, like in the case of the former.

Atomic pair distribution function (PDF) analysis, especially when combined with complementary techniques, is a highly suitable technique to tackle this so-called nanostructure problem.[2,3]

Atomic PDF analysis (described in detail in Chapter 2), is an experimental technique used to describe the local structure of samples, by employing a Fourier Transform to make full use of a diffraction pattern. This is accomplished through the calculation of a mathematical function called the PDF, which describes the distribution of distances between pairs of objects - in this context: pairs of atoms.

A close relative of the PDF is the radial distribution function (RDF), which describes the distribution of objects in space as a function of radial distance. It contains information regarding the shape, size, and local environment of the objects.

The primary focus of this thesis is the application of PDF and RDF analysis to the characterisation of the local structure in materials.

This was accomplished in three distinct manners, through the:

- Development of an image-recognition-based software tool called *ImageDataExtractor*, which auto-extracts and analyses imaging data from the literature to calculate various properties of detected particles, including their RDF.
- Development of a novel experimental technique called glancing-angle PDF (gaPDF), which allows PDF analysis of samples in the form of thin films and the application of this technique to working electrodes of dye-sensitised solar cells (DSSCs).
- Applications of PDF analysis to two materials case studies: the Gd-doped rare earth phosphate glass $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$, and four Ru-based photoisomers.

1.1 ImageDataExtractor

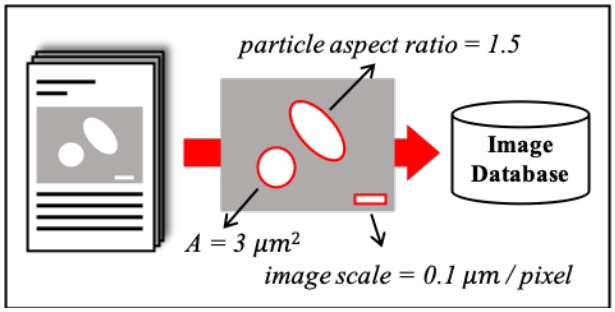


Figure 1.1 *An illustration of the auto-extraction and analysis process of images from the scientific literature performed by ImageDataExtractor.*

Data-driven materials discovery carries the idea that one can source and mine scientific data to establish structure-property relationships and uncover new scientific correlations. These new correlations can then be used to discover or re-purpose chemicals for a given materials application. To date, a large number of data-driven materials predictions have been presented using data generated by high-throughput computation.[4-8]

A significant hurdle is that the sought-after scientific data exist in a fragmented form across scientific papers and technical documents. Text mining tools relying on natural language processing and optical character recognition have been developed to partially overcome this challenge,[9-12] with a tool geared specifically towards the specialised language and textual-structure of the materials science domain being developed in the form of *ImageDataExtractor*. [13]

Despite these successes, tools that capture and quantify image-based data in this domain are starkly lacking, while images from microscopy techniques, containing a wealth of information, remain pervasive in scientific publications. Such images are often not quantified in the academic literature, in which cases the authors generally opt to discuss relevant features qualitatively, drawing quantitative conclusions about the material structure from less visual metrology techniques.

Computer vision, machine learning and Bayesian statistics have assisted the manual analysis of images from these specialised microscopy experiments to address very specific problems in materials science. For example, Holm and co-workers have used convolutional neural nets to analyse grain boundaries in SEM images of ultrahigh carbon steel.[14] Azimi *et al.* have classified microstructural features of steel in SEM and light optical microscopy images using fully convolutional neural networks.[15] Gola *et al.* used a support vector machine to classify microstructural images of low carbon steels.[16] Deep-learning methods have also been applied to STEM images in order to determine which crystal lattice defects are responsible for solid-state structural transformations in tungsten sulphide, [17] or semantic segmentation of defects in steels. [18] Bayesian inference has been used to extract interatomic interaction parameters of materials in STM images. [19]

ImageDataExtractor seeks to fulfil the data-mining need for a generic tool that auto-extracts and quantifies microscopy images from scientific documents, while harnessing the essence of these types of computing and mathematical capabilities that are being used for bespoke, manual, image quantification. It is capable of automated, high-throughput image extraction and quantification, providing information about the shape, size, and distribution of the 2-D projections of particles held within an image. The development of *ImageDataExtractor* is a significant first step toward the ultimate goal of auto-generating materials databases of information on particles from imaging data.

1.2 Glancing-angle PDF Analysis of DSSC Working Electrodes

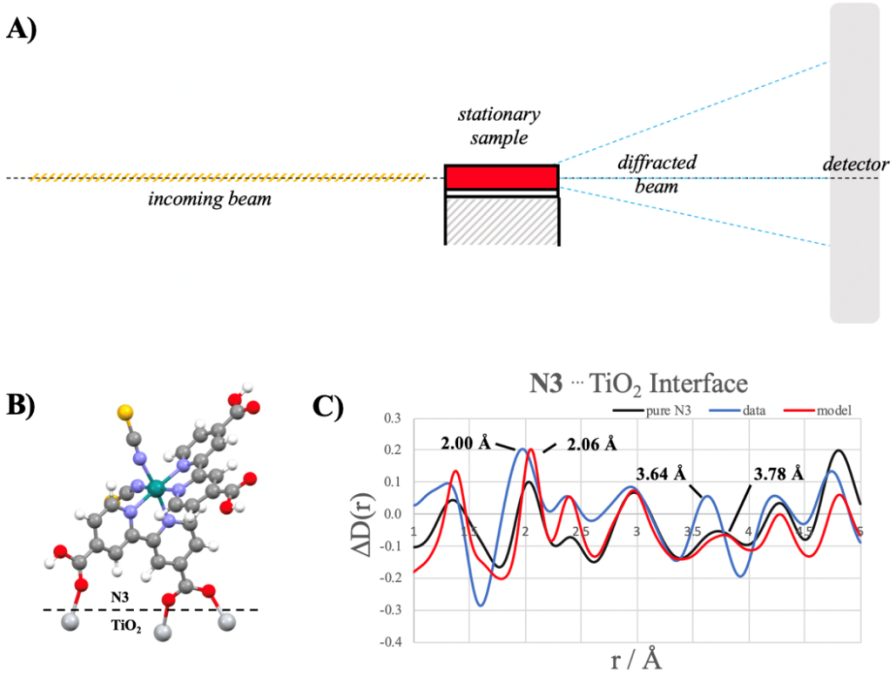


Figure 1.2 *(A) Schematic drawing of the experimental setup of glancing-angle PDF (gaPDF). (B) A model resulting from density functional theory (DFT) calculations on the dye* ***N3*** *adsorbed onto* TiO_2. *(C) The PDF agreement between the gaPDF data and model of* ***N3*** *on* TiO_2.

Dye-sensitised solar cells (DSSCs) represent promising prospects for the next generation of photovoltaic devices. [20] In particular, their transparency stimulates their development in solar-powered windows for energy-sustainable buildings in future cities.[21,22] In addition to representing the only truly transparent building-integrated photovoltaic technology, DSSCs excel in poor or ambient sunlight conditions. This stands in sharp contrast to other photovoltaic devices.[23,24]

The working electrode of a DSSC device usually adopts the form of a dye···semiconductor composite thin film, whose thickness is typically 5-10 µm, depending on the manufacturing method. The sunlight-harvesting dye molecules are bound to the semiconductor

(usually TiO_2) surface via an anchoring group, which is fundamental to the function of the DSSC since it facilitates the transfer of electrons from the photo-excited dye, initiating the electrical circuit. Therefore, the structural determination of a dye···TiO_2 interface, including a resolution of its preferred anchoring mode, is essential for the strategic improvement of DSSC technology.[25-27]

The structural nature of dye···TiO_2 interfaces has already been investigated using various materials characterisation methods such as: vibrational spectroscopy (IR and Raman), [28-31] X-ray photoelectron spectroscopy (XPS),[7,32,33] X-ray absorption spectroscopy, [33] optical metrology, such as UV-vis absorption and emission spectroscopy,[31,34] X-ray (XRR) and neutron reflectometry (NR), [8,35-39] concerted scanning tunnelling microscopy and scanning tunnelling spectroscopy (STM/STS),[40,41] and atomic force microscopy (AFM). [8,42,43]

These materials characterisation efforts have afforded discussions and conclusions in areas such as the ligands responsible for dye anchoring,[33] the existence of multiple binding modes[29,31,40] and variations in binding geometries.[28,32] However, the DSSC field still lacks a consistent, quantitative and reliable method to probe the local atomic structure of dyes on TiO_2.

PDF analysis has become a powerful technique for the characterisation of nanostructures,[1] having been borne out of a field of materials characterisation on liquids[44-47] and glasses,[44,48,49] which exclusively exhibit a local short-range order of the atomic structure.

PDF analysis has previously been applied to bulk powder (prepared by scraping off *meso*-$MAPbI_3$ from thin film electrodes) constituents of perovskite solar cell working electrodes, which revealed a two-component nanostructure, underscoring the need for a full characterisation in order to systematically improve cell performance. [50]

PDF analysis has also been successfully used to investigate the structure of thin films of $FeSb_3$, using a technique called thin-film PDF (tfPDF). [51] However, PDF analysis has not yet been used for the investigation of the thin-film working electrodes of DSSCs.

This study presents a novel experimental setup for PDF analysis, that we have termed glancing-angle PDF (gaPDF), which aims to analyse thin films of intact DSSC working electrodes. The study also employs complementary density functional theory (DFT) calculations to provide structural models, used to simulate corresponding PDFs of the two investigated interfaces, **N3**···TiO_2 and **N749**···TiO_2. The study also used data collected as a reference on pure **N3 dye** using a traditional PDF experiment setup. The chemical formulae of the dyes used are *cis*-bis(isothiocyanato)bis(2,2′-bipyridyl-4,4′-dicarboxylato)ruthenium(II) (**N3**) and tris(*N*,*N*,*N*-tributyl-1-butanaminium) [[2,2″6′,2″-terpyridine]-4,4′,4″-tricarboxylato(3-)*N*1,*N*1′,*N*1″]tris(thiocyanato-N) hydrogen ruthenate (**N749**).

Moderate agreement observed with the calculated models confirmed our ability to observe this interfacial structure using the gaPDF method. Deviations from the models, which employed one instance each of two separate binding modes (bidentate bridging and monodentate ester), showed shorter-than-expected bond lengths associated with the anchoring groups, suggesting a preference for the bidentate bridging mode, which features shorter overall bond lengths when compared to the alternative.

1.3 Gd···Gd Separations in Amorphous $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$

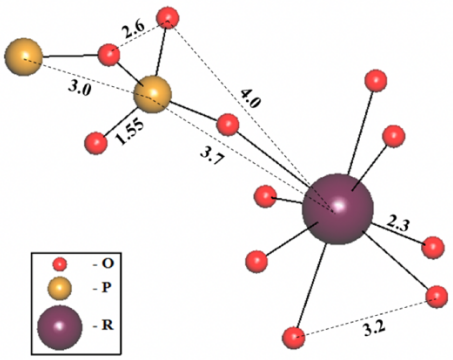


Figure 1.3 *Short-range order of rare-earth (R) phosphate glasses, $(R_2O_3)_x(P_2O_5)_{1-x}$, (x = 0.167 (ultraphosphate) to x = 0.25 (metaphosphate) showing its modelled geometry up to a radial distribution of r = 3.8 Å. Bond lengths presented in units of angstroms.*

Rare-earth phosphate glasses (REPGs) show great promise for applications in the laser and optoelectronics industry,[52-56] whereby REPGs of the type $(R_2O_3)_x(P_2O_5)_{1-x}$ (x = 0.167 (ultraphosphate) to x = 0.25 (metaphosphate)) are particularly interesting. REPGs in this stoichiometric range exhibit various attractive properties at low temperatures (T < 20 K), including acoustic mode softening, which results in negative thermal expansion coefficients [57] and bulk moduli,[58] as well as unprecedented magnetic, optomagnetic, and optoacoustic phenomena.[59]

A proper understanding the structure-property relationships of REPGs would help to significantly advance research and development in this area. The atomic structures of $(R_2O_3)_x(P_2O_5)_{1-x}$ (x = 0.167−0.250; R = La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, and Er) have been probed by a variety of spectroscopy and diffraction techniques, including conventional X-ray[60-67]and neutron[61,63,65,68] diffraction, as well as rare-earth K-edge [69] and L^{III}-edge [60,70-72] X-ray absorption spectroscopy (XAS), ^{27}Al and ^{31}P solid-state NMR,[62,68] and Fourier-transform IR spectroscopy (FTIR).[73] The combination of results from these previous studies has afforded a model of the local atomic structure of such REPGs, comprehensive only out to r ~ 3.8 Å.

However, the closest R···R separations cannot be discerned by these conventional techniques, implying this separation must lie at r > 3.8 Å. This structural feature is particularly important to define, given that rare-earth ion clustering can inhibit certain optical properties through quenching.

To that end, a Gd K-edge anomalous X-ray scattering (AXS) study was performed on the REPG $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$ in order to determine Gd···Gd separations in its local structure. My contributions to this study were the calculations and the statistical analysis of the ΔPDFs, which allowed drawing the structural conclusions.

A weak Gd···Gd pairwise correlation was located at 4.2(1) Å, which is representative of a *meta*-phosphate R···R separation. More intense first-neighbour Gd···Gd pairwise correlations were observed at larger radial distributions of 4.8(1) Å, 5.1(1) Å, and 5.4(1) Å. The local structure of this glass thus displays multiple rare-earth ion environments, presumably because its composition lies between the two stoichiometric formulae, $(R_2O_3)_x(P_2O_5)_{1-x}$ (x = 0.167 (ultraphosphate) to x = 0.25 (metaphosphate)).

To the best of our knowledge, this is the first report of a Gd K-edge AXS study on an amorphous material. This study demonstrates the ability to characterise the local structure of a REPG, heralding exciting potential prospects for AXS studies on other ternary non-crystalline materials. It also shows that AXS results are not only tractable, but they also reveal local structure of REPGs that is important from a materials-centred perspective and which could not be obtained by other materials characterisation methods.

1.4 Local Structure of SO_2 Photo-isomers in Ruthenium Coordination Complexes

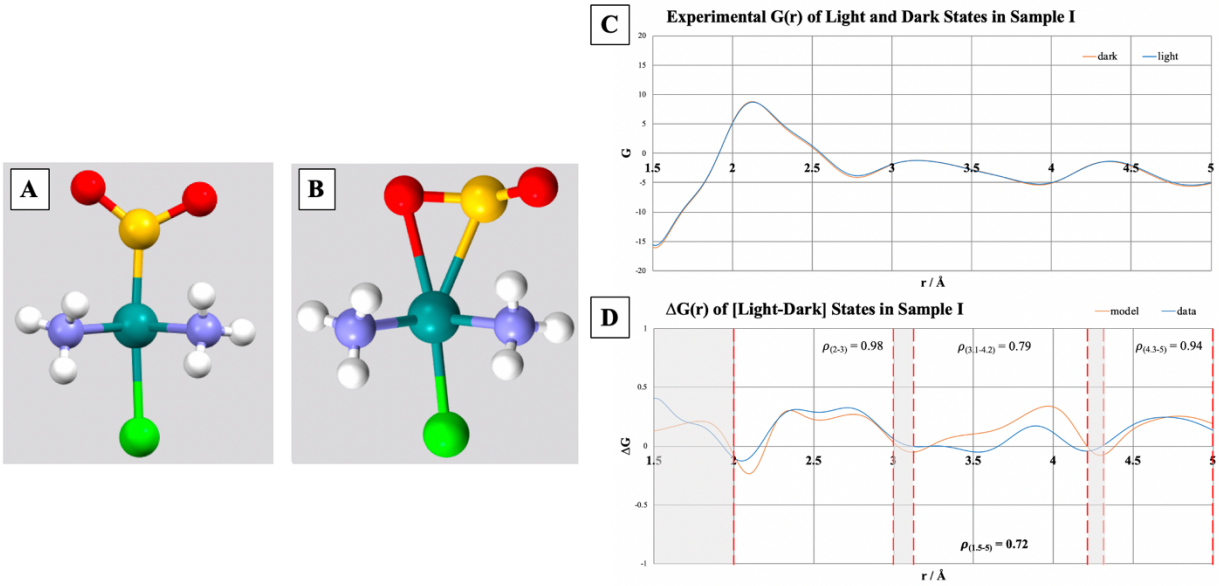


Figure 1.4 *Molecular structure of the cation in $[Ru(NH_3)_4SO_2Cl]Cl$ with SO_2 ligand in its (A) dark state (GS) and (B) photo-induced metastable state (MS2). (C) PDF analysis on data collected before and after photo-irradiation. (D) ΔPDF representing light-induced structural changes, compared to analogous ΔPDF calculated from photo-crystallography results.*

Materials that exhibit linkage photo-isomerism have a wide range of applications in the field of optoelectronics, given their optical switching[74] and nano-optomechanical transduction[75] characteristics. The optically accessible, metastable energy levels can allow a structure to function as a binary switch, with the ground state representing '0' and the excited '1', enabling use as logic gates or in 3D data storage. [76,77] Additionally, the steric consequences of the photo-activation process can be so severe that a secondary phenomenon (nano-mechanical transduction) is triggered, like in the case of a series of ruthenium-sulfur dioxide complexes,[78] where the photo-isomerisation of the SO_2 ligand in the cation of a coordination complex causes the neighbouring anion to move away from it.

More generally, a wider series of ruthenium-sulfur dioxide complexes have been found to behave as single-crystal optical actuators, exhibiting optical switching[79-85] or nano-optomechanical transduction[78,86] in their single-crystal form. The generic formula of these complexes is $[Ru(SO_2)(NH_3)_4X]Y$, where X is the ligand that lies in the *trans* position to the photo-active SO_2 ligand. Up to three different SO_2 configurations can co-exist within a structure in different fractions: the η^1-S-bound SO_2 dark state ('GS'), the most thermally stable photo-isomer, an η^2-(O/S)-side-bound SO_2 photo-isomer ('MS2') and the η^1-O-bound SO_2 photo-isomer ('MS1').

In situ light-induced single-crystal X-ray diffraction, known as photo-crystallography,[87-90] has been the primary method used to characterise the photo-induced crystal structures of these complexes, in attempts to understand the molecular origins of their linkage photo-isomerism. However, this method does not appear to tell the whole story. Correlated effects between light-induced atomic units, such as subtle levels of X-ray diffuse scattering or splitting of Bragg peaks, are often indicated in such studies.

The goal of this study is to investigate the local atomic structure of such materials, with a specific focus on light-induced changes, using PDF analysis of synchrotron scattering data, collected before and after sample photo-irradiation. $\Delta G_{exp}(r)_{light-dark}$ profiles were calculated in order to isolate the structural changes associated with photo-isomerisation. These were then compared to analogous $\Delta G_{sim}(r)_{light-dark}$ simulated from models constructed using photo-crystallography results.

Qualitative analysis of the structural features in the $\Delta G_{exp}(r)_{light-dark}$ was used to evidence the existence of local atomic structure in the samples as well as to demonstrate that photo-isomerisation is achievable in these compounds in their polycrystalline form; even when single-crystal samples are crushed into powders.

Visual peak analysis of $\Delta G_{exp}(r)_{light-dark}$ showed features consistent with the light-induced structural changes inferred from reference photo-crystallography results. Statistical comparison with analogous $\Delta G_{sim}(r)_{light-dark}$ calculated using models constructed from said photo-crystallography results showed high Pearson correlation coefficients (**I:** 0.98, 0.79, 0.94; **II**: 0.94, 0.84, 0.49; **III**: 0.98, 0.95, 0.26; **IV**: 0.86, 0.69, 0.67) over sample-dependent subdivisions of the r = [1.5-5] Å range, in decreasing order of association with light-induced structural changes. Disagreements in the $\Delta G(r)_{light-dark}$, particularly in the highest-r regions mentioned above, could possibly be caused by the unavoidable omission of 'mixed' type inter-ionic interactions (GS-MS1, GS-MS2, MS1-MS2), as well as other short-range order from the constructed models. This would point to the prevalence of such interactions, implying a lack of heterogeneity in the photo-isomerisation and, in turn, short-range order in these compounds. This study outlines an experimental procedure that paves the way for developing structural investigations on these types of materials using pair distribution function analysis.

2. Methods

2.1 Overview of Diffractive Techniques

Diffractive techniques rely on the interaction of incident radiation with a sample. The resulting diffraction pattern depends on both the crystallographic structure of the sample and its atomic composition. It is therefore possible to back-track information regarding the specific structural traits of a sample.

Traditional X-ray diffraction methods, developed more than 100 years ago, analyse Bragg peaks, *i.e.*, peaks in the diffraction pattern that are due to reflections from the most ubiquitous interatomic spacings present in the sample, ignoring the remainder of the pattern that is referred to as diffuse scattering. These methods provide information that agrees well with theoretical models, describing the average long-range order (LRO), while deviations of the short-range order (SRO) or inhomogeneous arrangements below a certain threshold are ignored.[91]

While these techniques dominated the field of diffraction analysis for more than a century, contemporary advanced materials require an in-depth understanding of their nanostructure, as it is often SRO that generates unique functionality. Nanoscopic examinations of chemicals with this short-range deviation from the average structure will manifest themselves as diffuse scattering. Since such chemicals have become of pivotal importance for the design and development of advanced materials, techniques that can provide insight into nanoscopic SRO have become indispensable materials characterisation tools.[92] Figure 2.1 (left) provides an example of a hypothetical material whose structure exhibits both SRO and LRO. Techniques describing average structure, such as traditional X-ray diffraction, would be sufficient to characterise the structure of this material, since its interatomic spacings are quite regular. The hypothetical structure in Figure 2.1 (right) exhibits only SRO. Therefore, only methods probing local atomic structure would be able to appropriately characterise the structure of this material.

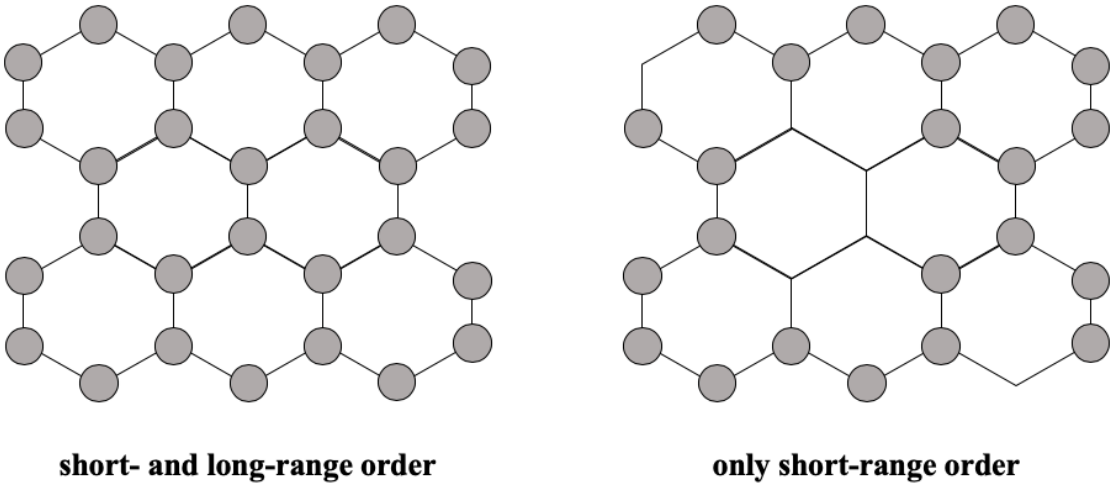


Figure 2.1 *Schematic of hypothetical materials exhibiting different kinds of structural order.*

Atomic PDF analysis, described in detail below, uses a Fourier Transform to make full use of the total scattering pattern, *i.e.*, Bragg peaks and diffuse scattering. This allows the description of both LRO and SRO, making it suitable to analyse structurally complex materials.

2.2 The PDF Method

The atomic PDF analysis method is an apt tool to uncover the information that is targeted in this project. It uses powder diffraction patterns to provide information regarding the local structure of a sample within the context of the average structure.

Although the roots of this method lie in experiments carried out almost 100 years ago, the potential of real-space resolution afforded by PDF analysis was not demonstrated until the late 1990s, thanks to advances in high-energy X-rays.[1]

This technique begins with the scattering vector, Q, the raw diffraction intensity, I_{exp}, the interference function, $i(Q)$, and the atomic scattering factor, f.

$$Q = 4\pi \frac{\sin\theta}{\lambda} \tag{2.1}$$

$$i(Q) = \frac{I_{exp} - \langle f^2 \rangle}{\langle f \rangle^2} \tag{2.2}$$

The atomic PDF can be defined by many inter-related functions, each of which is used for slightly different analyses of the same basic information, such as the total radial distribution function, $G(r)$, the differential distribution function, $D(r)$, or the total correlation function, $T(r)$, defined as:

$$G(r) = \frac{1}{(2\pi)^3} \int_0^\infty 4\pi Q^2 i(Q) \frac{\sin(Qr)}{Qr} dQ \tag{2.3}$$

$$D(r) = 4\pi r \rho_0 G(r) \tag{2.4}$$

$$D(r) = 4\pi r \rho_0 [G(r) + (\sum_{i=1}^{n} c_i \langle f_i \rangle)^2] \tag{2.5}$$

where c_i denotes the proportion of species, i, in the material and ρ_0 denotes the average atomic number density.

Each of these functions have corresponding partial functions, $g(r)$, $d(r)$, and $t(r)$, which represent contributions from individual pairwise correlations and have analogous definitions to the aforementioned total distribution functions.

Details of the atomic PDF analysis method and nomenclature, including data corrections and normalisations can be found in references [93] and [94].

2.3 Applications of the PDF Method

In practice, PDF analysis begins with a diffraction pattern. Figure 2.2 shows an example of a 2-D diffraction pattern, using data collected on sample **I** in Chapter 6.

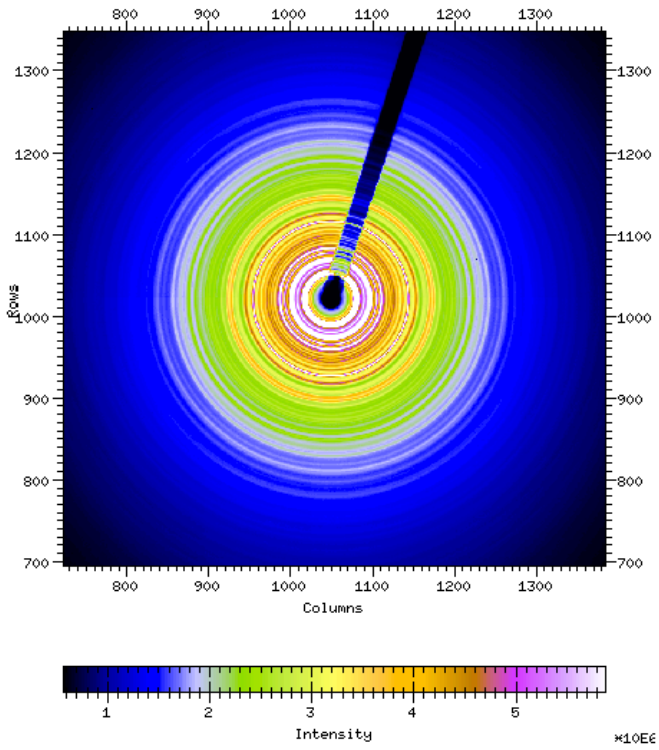


Figure 2.2 *A representative 2-D powder diffraction pattern obtained from sample* ***I*** *in Chapter 6.*

The sum of the 1-D cross sections of the Debye-Scherrer rings, which make up the 2-D pattern in Figure 2.2, afford I_{exp} (Figure 2.3).

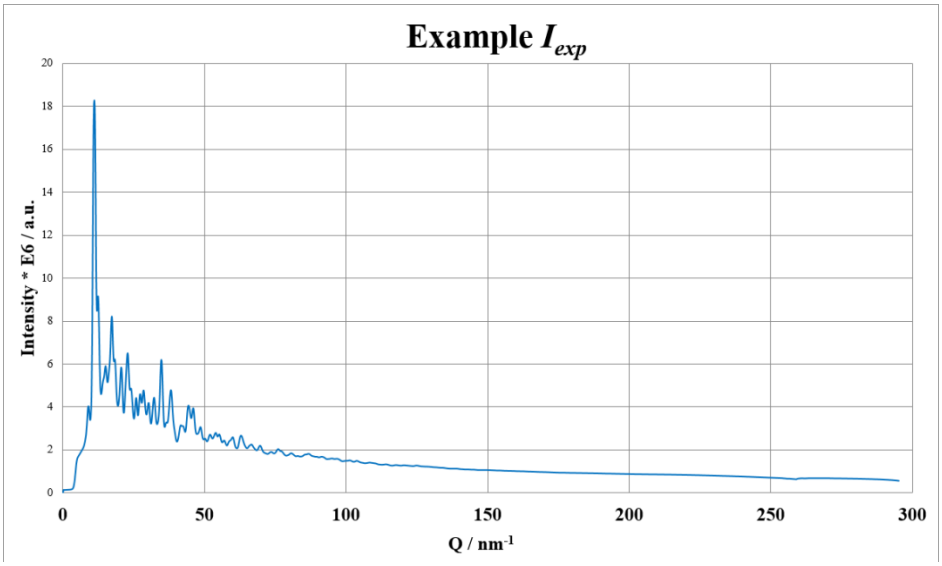


Figure 2.3 *A representative I_{exp} pattern obtained from sample* ***I*** *in Chapter 6.*

Equations 2.2 and 2.3 furnish the PDF, in this case: *G(r)* (Figure 2.4).

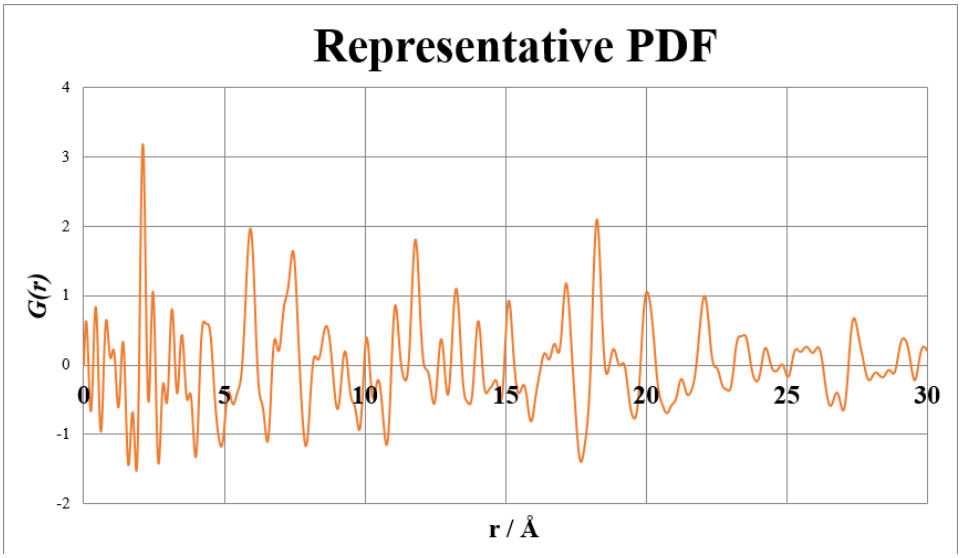


Figure 2.4 *Representative PDF, G(r), obtained from the diffraction data of sample* ***I*** *in Chapter 6.*

2.4 Analysing a PDF

A PDF represents the real-space scattering interactions between all pairs of atoms in the sample. Each pair-wise correlation contributes to the function on the *y*-axis (e.g., *G(r), D(r)*, or *T(r)*), at the radial distance, *r*, by which they are separated. The contributions from pairs of atoms at similar radial distances form a peak in the PDF, which is typically Gaussian in shape (Figure 2.5). The parameters of these peaks allow conclusions to be drawn about the SRO of the subject material.

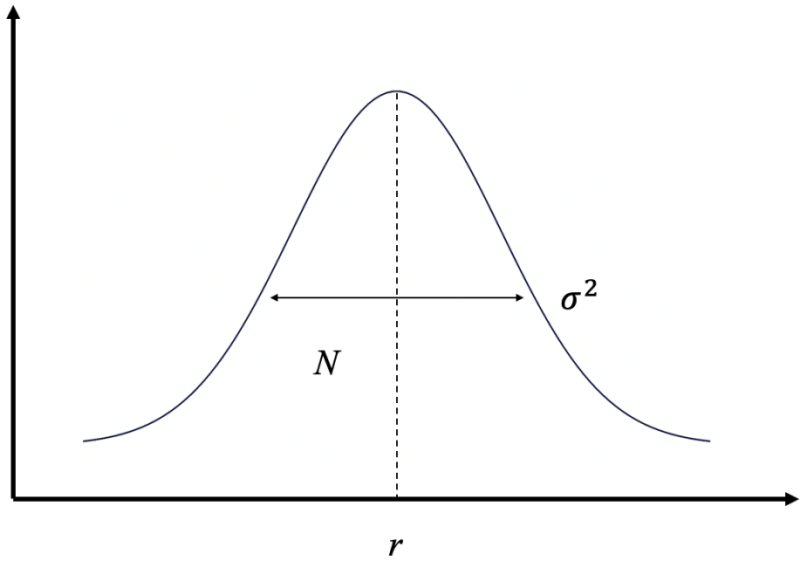


Figure 2.5 *Representative Gaussian peak of a PDF, together with the most salient peak parameters.*

The centre of the peak, *r,* denotes the average separation associated with a pairwise correlation. The area underneath the peak is related to the abundance of this pairwise correlation and can be used to determine its coordination number, *N*. The full width at half maximum (FWHM) of the peak provides the Debye-Waller factor, σ^2, which is a measure of the variation in this average radial separation. The FWHM is greatly influenced by temperature and disorder.

Samples that are less challenging in the context of PDF analysis are simple in both structure and composition, thus generating only few unique, first neighbour, pairwise correlations. These lead to well-defined Gaussian peaks that are easily traced back to its responsible atomic pairwise correlations.

More challenging samples will exhibit more complex structures and/or compositions which contain a multitude of unique atomic pairwise correlations. These lead to numerous, ill-defined, and usually overlapping peaks that are difficult to analyse.

Some samples that are simple in both structure and composition may still be challenging due to the nature of their atomic composition, which is a factor to be considered especially when using X-rays. Lighter atoms scatter X-rays more weakly than heavier atoms, making it easier for their contributions to the PDF to be dominated by background artefacts, noise, or those of heavier atoms that may also be present in the sample.

An example of a less challenging sample in the context of an X-ray-based experiment would be crystalline Ni, which is often used for calibration, since it strongly scatters and exhibits few types of first-neighbour pairwise correlations. An example of a more challenging sample would be a complex organic molecule, as it scatters weakly and exhibits multiple types of first-neighbour interactions, which leads to overlapping and ill-defined peaks.

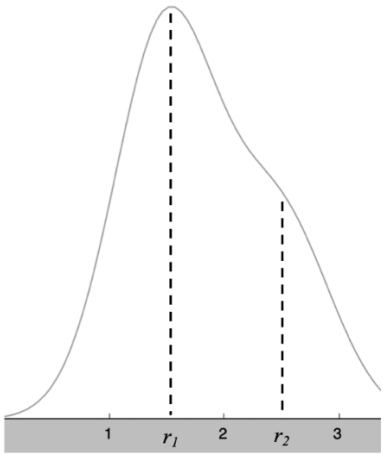


Figure 2.6 *An example of two overlapping Gaussian peaks, whereby the first is centred at r_1 = 1.5, while the second is centred at r_2 = 2.5; the amplitude of the peak at r_2 is half that of the peak at r_1.*

Figure 2.6 shows an example of two overlapping Gaussian peaks, whereby the first is centred at r_1 = 1.5, while the second is centred at r_2 = 2.5. The second, smaller, peak is often referred to as a "shoulder" of the first peak. In this example, the smaller peak has half the amplitude of the first peak. If the amplitude of the second peak were to become a bit smaller, the shoulder would be too subtle to discern, especially once experimental noise is considered. An example of such an instance might be the reduction in compositional stoichiometry of a material, whereby the pairwise atomic correlation responsible for this composition pertains to this shoulder. Such cases are ubiquitous during the analysis of challenging samples.

2.5 Difference PDFs

The analysis of such subtle features may benefit from more specialised sub-methods within PDF analysis. Often, such features are subtle because they concern only a particular part of the structure of a material rather than the entire structure. They are dominated by other unambiguous, but irrelevant, features. Therefore, such analysis calls for techniques that allow the isolation of specific pairwise correlations that are responsible for the features in question.

Difference PDFs (ΔPDFs) allow a subtraction of unwanted X-ray scattering contributions that originate from the irrelevant parts of the structure, thus isolating the specific structural feature that is targeted. This is accomplished by repeating the experimental process twice on a sample whose X-ray diffraction patterns only differ by the structural feature in question. The corresponding PDFs can then be subtracted, which isolates the PDF signal that describes only the structural difference.

Many of the conclusions presented in Chapters 4-6 are based on the ΔPDF technique.

One of the samples analysed in Chapter 6, *trans*-[Ru(NH_3)$_4$(SO_2)Cl]Cl (sample **I**), is now provided as an example. In summary, this study investigated the light-induced structural changes in four compounds, including sample **I**, by applying ΔPDF analysis to synchrotron-based X-ray powder diffraction data collected before and after sample photo-irradiation.

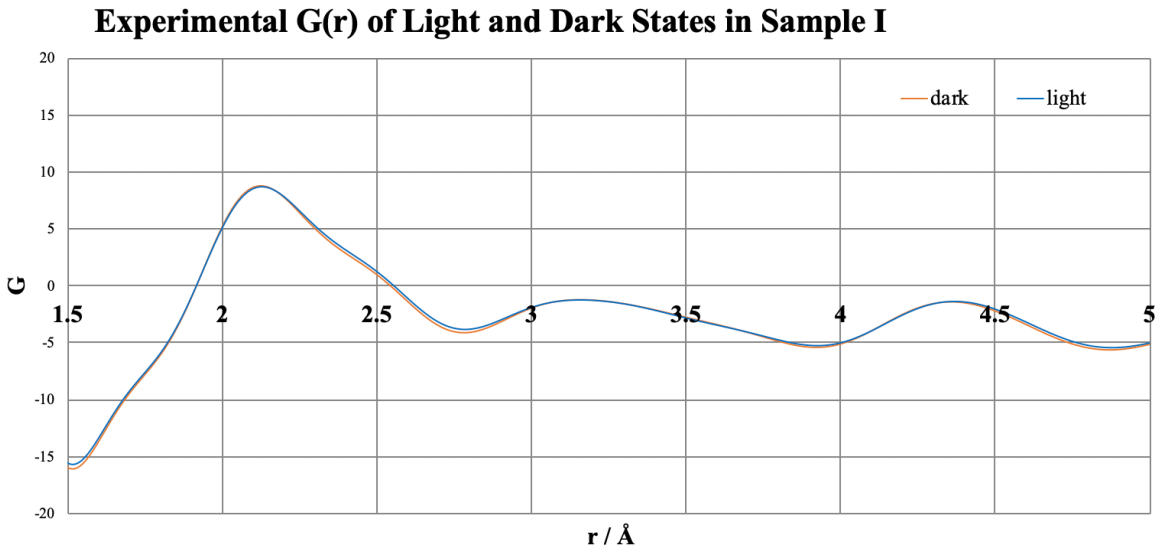


Figure 2.7 *G(r) of sample I, collected before (dark state, orange) and after (light state, blue) photo-irradiation.*

As can be seen in Figure 2.7, the *G(r)* profiles of the sample in its dark and light-induced states are exceptionally similar, showing a Pearson correlation coefficient of ρ = 0.99964 over the range r = 1.5 – 5 Å, where a value of 1 would indicate a perfect correlation.[95] It would be very challenging to try to draw conclusions regarding the light-induced structural changes of this sample using its *G(r)* data in this form.

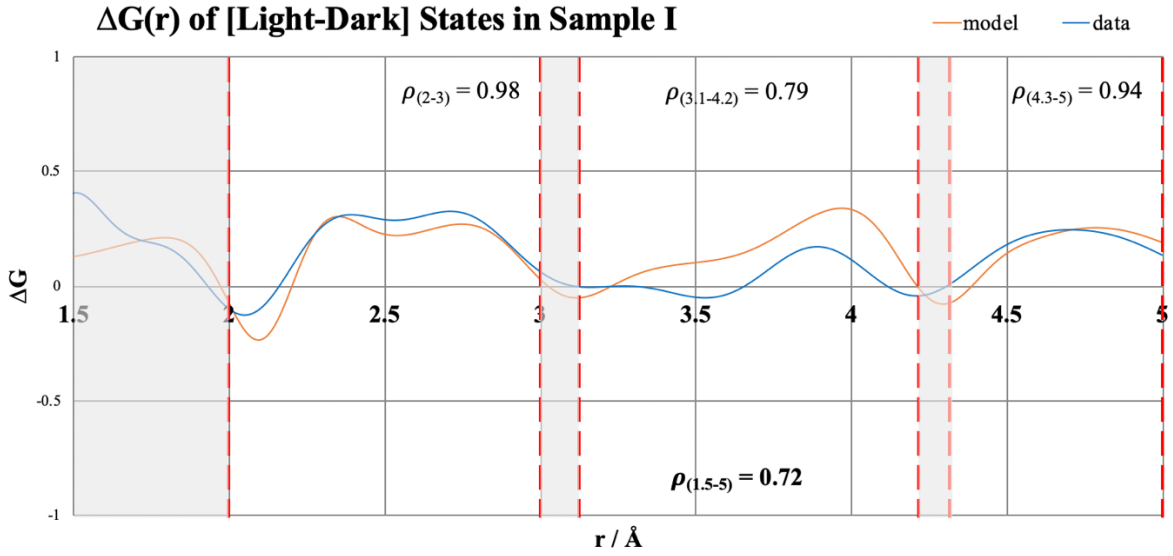


Figure 2.8 $\Delta G(r)_{light-dark}$ *for sample* ***I*** *generated from experimental data (blue) and from simulated models (orange) that account for its light-induced structural changes.*

Figure 2.8 shows the $\Delta G(r)_{light-dark}$ for sample **I**. The difference calculated from the data presented in Figure 2.7 is shown in blue, presenting the peaks that manifest light-induced structural changes much more clearly than in the previous step. A comparison of this experimental difference with analogous Δ*G(r)* profiles, calculated from structural models from photo-crystallography, shows strong statistical agreement, as can be seen in the high Pearson correlation coefficients annotated in Figure 2.8; this strongly supports their validity. Pearson coefficients above 0.70 were considered indicators of strong agreement in this context, given the inherent drawbacks of applying a photo-crystallographic model to ΔPDF data.

2.6 Statistical Methods

The example presented above made use of the Pearson correlation coefficient. This is a statistical coefficient, among others such as the Kendall or Spearman's rank,[95] that is often used on PDFs[96,97] and ΔPDFs.[98-100] It can be used to quantify an observed agreement or confirm the existence, or lack thereof, of structural similarities between separate cases. The assumptions of these correlation coefficients can vary slightly; for example, Pearson's coefficient will assume a linear relationship between two data sets while the latter two make no such assumption.[95] Regardless of their relative sensitives, such coefficients ignore absolute scaling, while remaining sensitive to relative peak scaling as well as shifts in peak positions, making them extremely suitable for comparing PDFs.

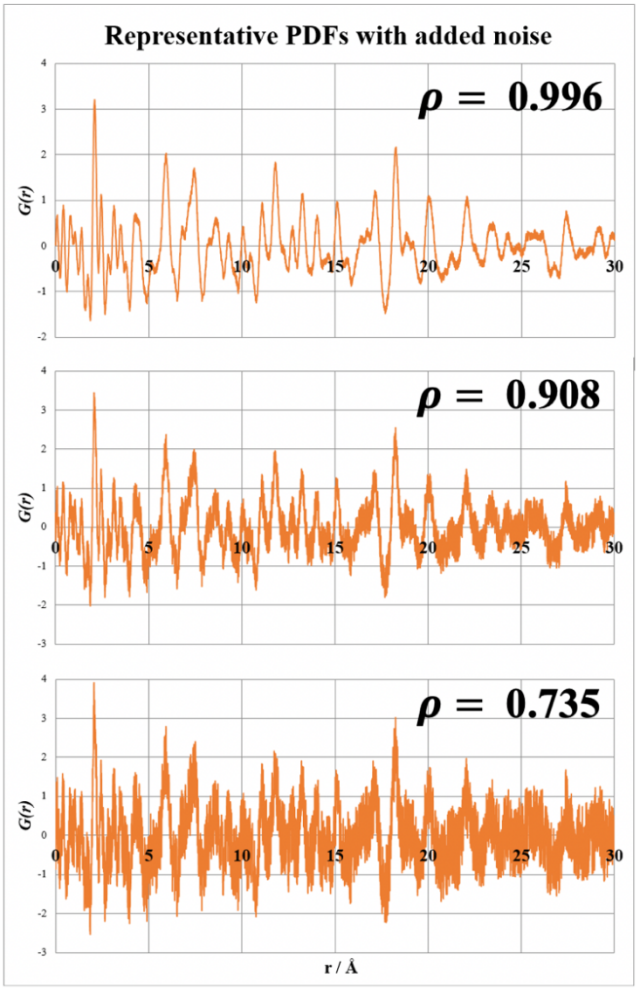


Figure 2.9 *Representative PDFs with increasing amounts of added noise from top to bottom. Annotations show Pearson coefficients between each PDF and the original from Figure 2.4.*

As can be seen in Figure 2.9, these statistical coefficients remain capable of identifying correlations, even in situations where data exhibit poor signal-to-noise ratios. Such statistical analysis helped draw many of the conclusions presented in Chapters 4-6.

3. ImageDataExtractor

The first instance of radial distribution function analysis is through an image-recognition based software, *ImageDataExtractor,* which extracts quantitative information, such as particle size and particle distribution function, from microscopy images. This chapter is based on a paper published in the *Journal of Chemical Information and Modeling*[101] (see *Publications* on page II). All figures and tables are reproduced in accordance with the guidelines set by the publisher. http://pubs.acs.org/articlesonrequest/AOR-bS5eBXSMECNNakJsmGi2

My PhD supervisor, Jacqueline Cole, and I conceived and designed the research together, working closely on developing the concepts associated with each stage of the work. I wrote the majority of the code. Batuhan Yildirim, also from the Molecular Engineering group, wrote the SRCNN module (explained in Chapter 3.3, Step 2). The code was prepared for installation and embedded into *ImageDataExtractor* by Edward Beard, who also created the web platform https://www.imagedataextractor.org/.

3.1 Summary

The rise of data science is leading to new paradigms in data-driven materials discovery. This carries an essential notion that large data sources containing chemical structure and property information can be mined in a fashion that detects and exploits structure-property relationships, such that chemicals can be predicted to suit a given material application. The success of material predictions is predicated on these large data sources of chemical structure and property information being suited to a target application. Microscopy is commonly used to characterise chemical structure, especially in fields such as nanotechnology where material properties are highly dependent on the size and shape of nanoparticles. Large data sources of nanoparticle information stemming from microscopy images would thus be highly beneficial. Millions of microscopy images exist but they lie fragmented across the literature, typically presented

individually within a paper article and usually in a qualitative fashion therein, even though they harbour a wealth of numeric information. We present the *ImageDataExtractor* toolkit that auto-identifies and auto-extracts microscopy images from scientific documents, whereupon it autonomously analyses each image to produce quantitative particle size and shape information about its subject material. Each image is quantified by decoding its scale bar information using optical character recognition, with help from super-resolution convolutional neural networks where required. Individual particles are detected and profiled using various thresholding, segmentation, polygon fitting, and edge correction routines. The high-throughput operational capability of *ImageDataExtractor* means that it can be used to generate large-data sources of particle information for data-driven materials discovery. Evaluation metrics, precision and recall, are greater than 80% for the majority of the image processing steps, and precision is above 80% for all critical steps. The *ImageDataExtractor* tool is released under the MIT license and is available to download from http://www.*imagedataextractor*.org.

3.2 Introduction

The rise of data science is heavily indebted to major advances in artificial intelligence (AI) that can probe large arrays of data to uncover patterns that refer to previously unseen scientific correlations. AI has a data dependency in return, often requiring massive data sets to realise its full potential. Large scientific datasets have started to become openly available via government-regulated open-access requirements for data and literature. Data science in the materials domain has been stimulated specifically by the US government via their Materials Genome Initiative; [102] this seeks to reduce the average 10-20 year 'molecule-to-market' industrialisation timeframe by finding innovative ways to discover and manufacture advanced materials for a given device application. Data-driven materials discovery has been forged from this initiative.

Data-driven materials discovery carries the idea that one can source and thence mine data about chemicals and their cognate material properties. The data-mining pursuit involves identifying relationships between chemical and material properties and then employing these structure-property relationships as knowledge-based rules or training data to find new data correlations. This leads to the discovery of new or re-purposed chemicals for a given materials application.

A large number of data-driven materials predictions have been presented using data generated by high-throughput computation.[4-6] Such data sources tend to be sparsely, if at all, benchmarked by experimental data, while the material predictions themselves are rarely experimentally validated. A notable exception is the data-driven materials discovery of organic light-emitting diodes [103] that was informed by computational data.

Successful examples of data-driven materials discovery that employ both experimental and computational data in the data-mining workflow are even rarer, but do exist in the realisation of new light-harvesting chromophores for photovoltaic applications.[7,8] The chromophores discovered by Cole *et al.* [7] used the Cambridge Structural Database [104] as the experimental data source while Cooper *et al.* [8] employed a UV/vis absorption spectroscopy database [105] for their materials discovery objectives. Given that structure-property relationships lie at the core foundation of data-driven materials discovery, materials characterisation data from metrology measurements offer the primary experimental sources of information on chemical structure for data-mining. Yet, there is a dearth of openly available materials databases of this nature. Accordingly, experimental databases are commonly unavailable to solve a data-mining task in the materials-science domain, even though the data themselves often exist in a fragmented form, e.g. in academic papers, or technical documents such as patents or company reports.

Text-mining tools have been developed in order to partially overcome this hurdle whereby fragmented data on chemical materials and properties are auto-extracted from documents and automatically stored in custom materials databases for a bespoke application.[9-12]

The text-processing capabilities of such tools rely on developments in natural language processing (NLP) and optical character recognition (OCR) that were borne out of computer science; yet, with adaptation in order to cope with the highly specialised language and textual-structure of the materials science domain. It was with this spirit that the 'chemistry-aware' text-mining tool, *ChemDataExtractor* [13] was developed. This tool has already been used successfully to auto-generate a database for metrology e.g. an open-source database of UV/vis absorption spectra [105] and a materials database for chemical properties, e.g. a materials database of magnetic Curie and Néel phase-transition temperatures. [106]

Despite these successes in auto-generating materials databases via text-mining, tools that capture and quantify image-based data in the materials domain are starkly lacking, at least at the generic level. Yet, images from microscopy techniques are pervasive in scientific publications, notably in areas of nanotechnology and crystallography, and they contain a wealth of information. This is particularly true of scanning electron microscopy (SEM), transmission electron microscopy (TEM) and atomic force microscopy (AFM) which act as the most common workhorse image-based materials characterisation methods in a laboratory. Such images shown in the academic literature are rarely quantified, although certain image features are often discussed qualitatively within the paper in which they feature. Moreover, quantitative conclusions about a material structure in a given study tend to come from one or more, less visual, metrology technique presented elsewhere in a scientific document.

Image quantification is nonetheless possible in microscopy, as is exemplified by research carried out on more specialised, low-throughput, microscopy techniques, such as scanning TEM (STEM), scanning tunnelling microscopy (STM) and scanning X-ray microscopy (SXM). The typical time and resource investments on these types of experiments are several orders of magnitude higher than that of a routine image-based materials characterisation measurement. Thus, the researcher is far more inclined to invest manual effort into image analysis to quantify their results. Moreover, a scientific problem that has been pre-selected for such a resource-

intensive experiment is *a priori* high priority, such that the images accrued deserve as quantitative an understanding wherever possible.

Computer vision, machine learning and Bayesian statistics have assisted the manual analysis of images from these specialised microscopy experiments, to address very specific problems in materials science. For example, Holm and co-workers have used convolutional neural nets to analyse grain boundaries in SEM images of ultrahigh carbon steel [14] Azimi *et al.* have classified microstructural features of steel in SEM and light optical microscopy images using fully convolutional neural networks. [15] *Gola et al.* used a support vector machine to classify microstructural images of low carbon steels.[16] Deep-learning methods have also been applied to STEM images to determine which crystal lattice defects are responsible for solid-state structural transformations in tungsten sulfide, [17] or semantic segmentation of defects in steels. [18] Bayesian inference has been used to extract interatomic interaction parameters of materials in STM images. [19]

This study seeks to fulfil the data-mining need for a generic tool that auto-extracts and quantifies microscopy images from scientific documents, while harnessing the essence of these types of computing and mathematical capabilities that are being used for bespoke, manual, image quantification. To this end, we present a new software tool, *ImageDataExtractor*, which is capable of automated, high-throughput, image extraction and quantification. We describe the three main components of the program: an image auto-extraction tool to mine images from documents such as the academic literature, an image recognition algorithm that identifies objects of interest, and an analysis suite that quantifies these objects where they are particles in a sample, in terms of shape, size and distribution. We evaluate the tool using standard precision and recall metrics and showcase examples of its application. *ImageDataExtractor* can be applied in one of two modes. It may operate as a stand-alone tool where it receives images directly, in various formats. Alternatively, it may function with *ChemDataExtractor* [13] as a plug-in, in cases where scientific documents are the input. Thereby, *ChemDataExtractor* [13] first acts to locate prospective microscopy images within documents by searching the text of figure captions for

keywords such as SEM, TEM or AFM; *ImageDataExtractor* then proceeds to assess if this is a true image and, pending this to be the case, extract and quantify the image. Our system overview below describes primarily the operational mode when *ChemDataExtractor* [13] is used as a plug-in, since this is a superset of the other, simpler, mode of function.

3.3 Implementation

System Overview

Figure 3.1 shows the system overview of *ImageDataExtractor* when in its operational mode that incorporates the full processing workflow, from the input of scientific documents to output that expresses features within images in a quantifiable manner. There are three essential steps in this pipeline: (1) image identification and auto-extraction from documents; (2) image recognition of material objects of interest; (3) quantitative analysis of material objects of interest in the images. Each step will now be described in detail via the following sections.

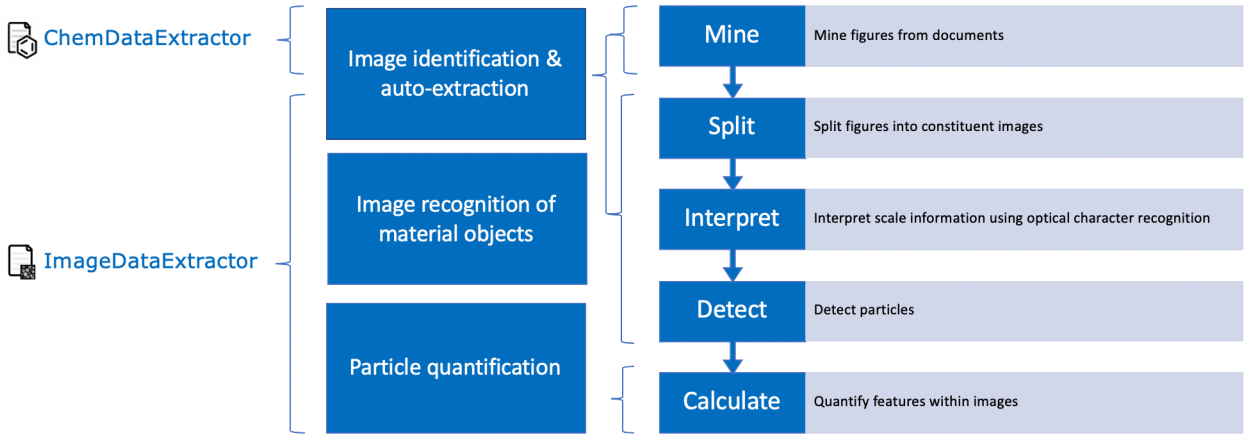


Figure 3.1 *Overarching system pipeline for ImageDataExtractor.*

Step 1: Image auto-identification and auto-extraction from documents

ChemDataExtractor [13] may be used as a prelude to *ImageDataExtractor*, in order to mine the text of figure caption which helps to identify prospective microscopy images that comprise, or reside within, figures of documents. *ImageDataExtractor* then mines these identified figures, splitting them into their constituent parts where required, to auto-extract images that pertain exclusively to microscopy. More detailed mining and splitting procedures are now described.

Mining Figures from Documents. One or more files, most conveniently formatted in mark-up language, HTML or XML, are used as input for *ChemDataExtractor*. [13] This software distinguishes the various structural components of the document; for example: headings, paragraphs, figures, tables. All figure objects are then extracted from each document together with a tokenised version of its figure caption and a URL link to its online source. The tokenised caption is then parsed using a set of simple rules to identify keywords that are indicative of a microscopy image, such as the acronyms 'SEM' and 'TEM'. When the conditions of these rules are met, the figure object is downloaded using its URL link and an associated text file is temporarily saved in CSV format; this is populated with metadata including the article's Digital Object Identifier (DOI), the figure's unique identifier (e.g. this is '4' for 'Figure 4' within the original document) and its caption text. For the purposes of illustrating the use of our new tool, the auto-extraction feature of *ChemDataExtractor* [13] for *ImageDataExtractor* input has been set up to support downloads from the Royal Society of Chemistry publishing service. However, it could in principle support image auto-extraction from any available document source.

Splitting Figures into Constituent Images. A given figure object may comprise exclusively a microscopy image. Yet, microscopy images are commonly placed together with other types of images to form an *image panel* that constitutes an overall figure within a document. Microscopy images present in such panels need separating from the other images in

the figure, in order that they can be analysed. An example of an image panel is given in Figure 3.2, together with a conceptual illustration of the need to separate the microscopy images.

The separation process is performed in *ImageDataExtractor*, where required, and this comprises two steps: firstly, binarisation, secondly, a bespoke method that examines an image for signatures of straight lines. The binarisation step is simply a means of converting image pixels into black and white to signify if they belong to an image or not. It is carried out with a very high threshold to ensure that the extractions represent a photographic image. The resultant black rectangles indicate a unique micrograph, which defines a set of bounding boxes for segmentation. The second step employs a bespoke figure splitting algorithm which splits any figures that use minimal, to no, borders while collating images of identical sizes. The algorithm searches for long, straight, axes that manifest along fractions of the width and height of the image. This is accomplished using the built-in OpenCV[23] function HoughLinesP, which implements the probabilistic Hough line transform, as described by Matas *et al.* [107] The nature of such lines determines if a figure should be split vertically or horizontally, and into how many parts (e.g. half, thirds, quarters).

ChemDataExtractor Plug-in versus Stand-alone ImageDataExtractor Operation. *ImageDataExtractor* supports high-throughput image extraction for large-scale analysis in either its stand-alone version or in its form where it is used in concert with *ChemDataExtractor.* [13] However, it can only achieve high-throughput extraction from documents when *ImageDataExtractor* operates with *ChemDataExtractor* [13] as a plug-in. Thereby, its input is a corpus of articles, which can be obtained in large quantities using the 'scrape' package provided inside *ChemDataExtractor* [13] to search for relevant articles from a range of supported publishers. Alternatively, users can access the Application Programming Interfaces (APIs) of specific publishers directly to create a corpus. *ChemDataExtractor* [13] processes these documents, identifying the images as well as producing metadata. The stand-alone version of *ImageDataExtractor* extracts one or more images directly from image file format, and thus does not produce any metadata.

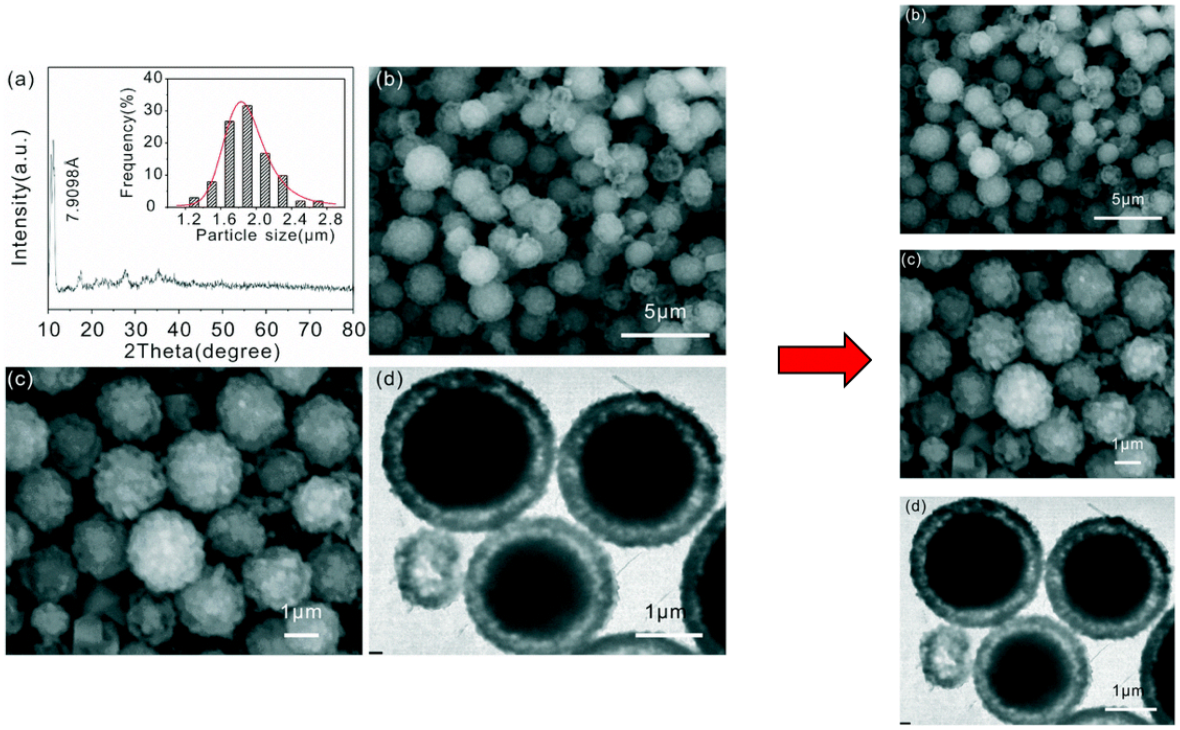


Figure 3.2 *The original figure extracted from an article (left), the three microscopy images extracted (right) from the original figure. Reproduced with permission from reference [108]. Please note the figure labels (a)-(d) were carried over from the original publication and can be disregarded for our purposes.*

Step 2: Image recognition algorithms that identify material objects of interest

This step can be broken down into two distinct sequences of manifold operations, one that determines the scale of a micrograph and another that detects individual particles within an image. Details of each operation are now given.

Interpreting scale information. This process begins with the identification and isolation of regions in an image where text is suspected to exist. Text usually appears white or black in a greyscale image, so text can be found by looking for image regions that feature very high or low grayscale pixel intensity. Once text in these regions has been identified, a convolutional neural

network is used to enhance textual resolution, where this is deemed necessary, to facilitate the next step which employs optical character recognition (OCR). All image text is read by OCR and stored. The scale bar in the image is then identified. Thereby, OCR-resolved text that contains values and units is distinguished, since such text is typically associated with a scale bar. The image region surrounding this scale-bar text is then searched for rectangular shapes that may denote a scale bar. Once located, three constituent parts of the scale bar (its width, textual value and unit) are drawn out from the image, in order that this scale bar can be used to calculate the scale of the actual image. This process is summarised in Figure 3.3, while the details are now discussed.

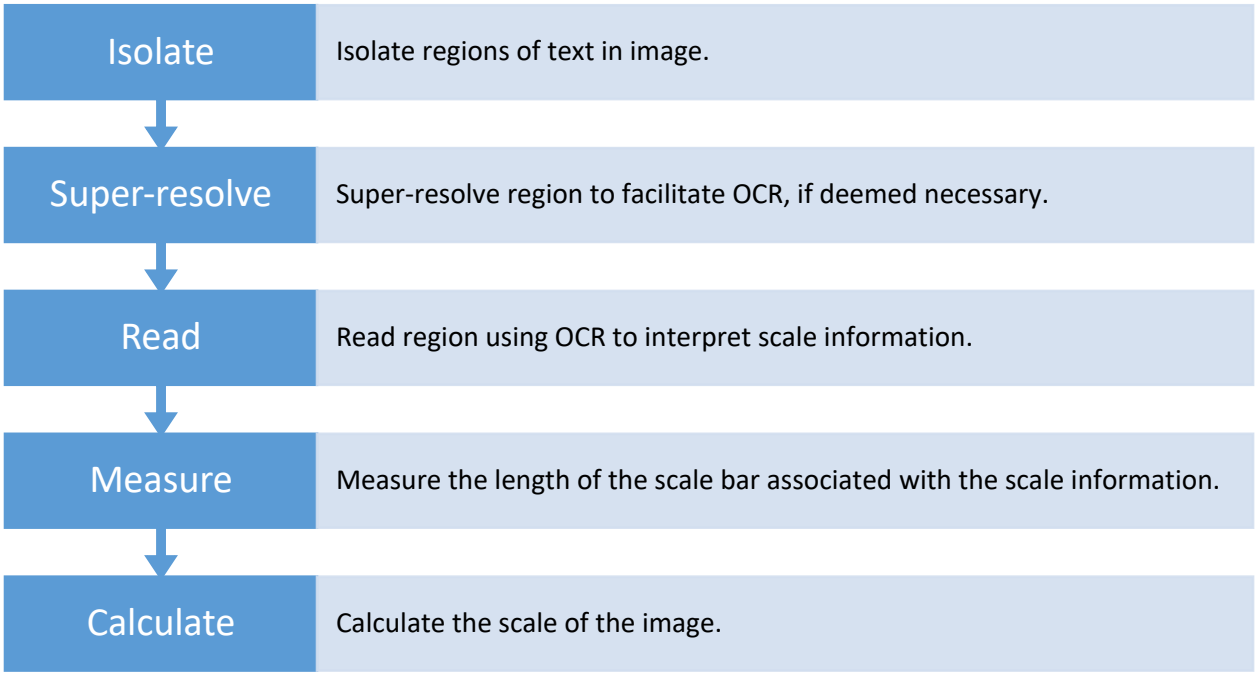


Figure 3.3 *Process for interpreting scale information in images.*

All image manipulations are carried out using built-in OpenCV [109] functions or manual array editing, unless otherwise noted. Details of the algorithms employed by the OpenCV functions are provided below where necessary.

Isolating Regions of Text. The isolation of text in images begins with the conversion of the image to grayscale and its subsequent binarisation, with a threshold value calculated according to:

$$t = \mu_{px} + 3.5 \times \sigma_{px} \tag{3.1}$$

where μ_{px} is the mean pixel intensity and σ_{px} is the standard deviation of the pixel intensity of the image. A ceiling value of 235 is applied. This process ensures that any pixel with an intensity lower than the threshold value is set to 0 (black) and all others to 255 (white), leaving behind white text and often small artefacts elsewhere in the image.

Next, contour detection is applied to the filtered image using the OpenCV function, findContours, which isolates the regions of an image that are of interest via the border detection algorithm described by Suzuki *et al*. [110] This algorithm considers a given candidate region in a filtered image and assesses if it includes an artificial box (example in Figure 3.4) upon which the text was placed, as opposed to text placed directly on the image. This invokes the contour detection method and a search for strictly rectangular shapes that include further shapes within them. If an artificial box is found, the candidate region is cropped to contain only its area, and then passed onto OCR processing. The corresponding regions from the unfiltered image, whose rectangular bounding boxes are larger than 1% of the total image area, are also passed on to OCR processing.

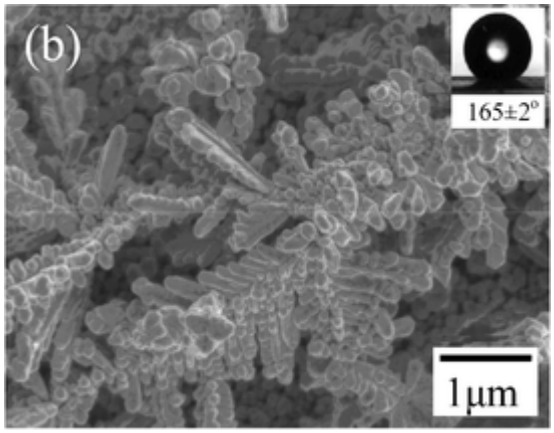


Figure 3.4 *An example of an image featuring scale information placed on an artificial box (bottom right).* [111] *Please note the figure label (b) was carried over from the original publication and can be disregarded for our purposes.*

Next, OCR is applied to the textual regions of interest, using the Python wrapper Pytesseract [112] of the open-source OCR engine, Tesseract 3. [113] Specifically, OCR is applied to the unfiltered region of interest, as well as seven different binarised versions of the colour-inverted region, since the OCR engine was found to perform better on black text on white background. Binarisation is applied using absolute threshold values (any pixel with an intensity above this value is set to white while all others are set to black), ranging from grayscale pixel intensities of [5,35] in increments of 5, until the text surveyed fitted several bespoke conditions. These conditions are that: (i) the text must include the phrase "nm" or "um" (it was found that the default Tesseract library performed better searching for "u" instead of "μ"), preceded either by digits or whitespace and then digits; (ii) the numeric value of the text must not be smaller than 0.01 or larger than 1000 and it must be evenly divisible by 5 if larger than 10. These conditions greatly reduce the risk of incorrect readings. Other units of length such as "cm" or "mm" were excluded from condition (i) since their appearances in the literature were not high enough to justify the reduced precision their inclusion introduced to this step, mostly due to confusions of "mm" with "nm".

If this OCR processing stage is unsuccessful, a convolutional neural network is applied to the candidate region to increase its resolution and the same OCR process mentioned above is then repeated.

Facilitating OCR using Super-Resolution Convolutional Neural Networks. The image resolution of the isolated text is then enhanced where required using a super-resolution convolutional neural network (SRCNN),[114,115] albeit in a modified form as described below. The SRCNN method first upsamples an input image x_0 with dimensions (H x W) by a scale factor, s, using bicubic interpolation. Its output is then passed to the convolutional neural network[116] as an input. A series of convolution operations, with non-linear mappings using the rectified linear unit (ReLU) operation,[117] are then performed. The output is a high-resolution image with sharper edges than the original image. The entire process is represented as:

$$F : \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{sH \times sW} \tag{3.2}$$

where F is the upsampling operator which produces the super-resolution image.

Our specific SRCNN is composed of three layers, each consisting of a depthwise separable convolution[118] and a ReLU non-linearity,[117] with the exception of the output layer which omits the ReLU operation. The first layer learns a set of feature maps from the input image, in the form of a set of high-dimensional vectors. The second layer is a 1 x 1 depthwise separable convolution in which no spatial information is used and the features are mapped from one high-dimensional channel space to another. Finally, the output layer aggregates the feature maps into the super-resolved pixels, to afford the high-resolution image as output.

We modified the original SRCNN architecture by replacing the traditional convolution operation with depthwise separable convolutions[118] to realise improved speed and performance. Depthwise separable convolutions consist of a spatial convolution performed on each channel of the input, followed by a 1 x 1 pointwise convolution which projects the output of the spatial

convolutions onto a different channel space. Thus, channel-wise features are learned separately from spatial features.

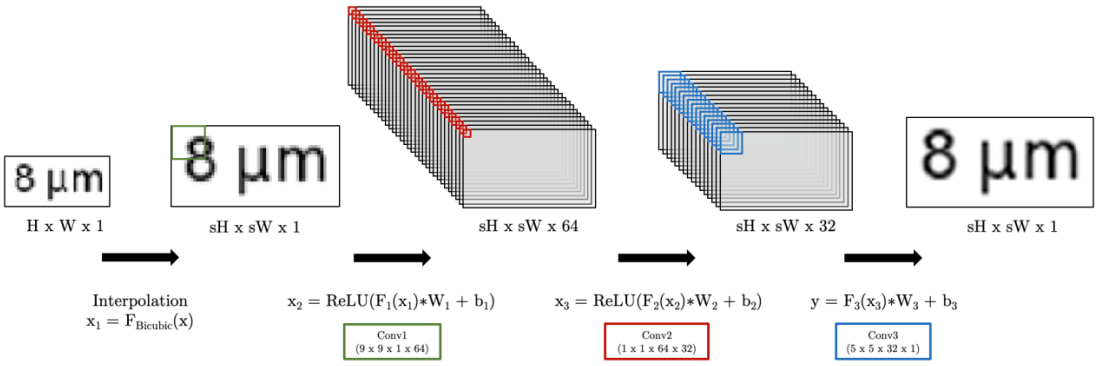


Figure 3.5 *The super-resolution module, consisting of four operations.*

Figure 3.5 describes the four operations of which make up this module: (i) The scalebar text image is resized by bicubic interpolation. (ii) The first convolution operation (∗) performs an aggregation of each pixel and its neighbours in a 9 x 9 region via convolution by a (9 x 9 x 1 x 64) weight matrix W_1, to produce a 64-dimensional feature vector for each pixel. This is followed by a ReLU non-linearity, which is applied element-wise to each pixel. (iii) The second convolutional layer produces for each input 64-dimensional pixel feature map, a subsequent 32-dimensional pixel feature map. In this step, convolution is performed by a (1 x 1 x 64 x 32) kernel weight matrix W_2, meaning pixel feature maps are not aggregated with their neighbours. This step also includes a ReLU non-linear transformation, which follows the convolution. (iv) The final convolution operation which produces the output super-resolution image, convolves each 32-dimensional feature map by a (5 x 5 x 32 x 1) weight matrix W_3. This is purely a linear transformation with no non-linearity, which maps each 32-dimensional pixel feature map from the previous layer, to a single super-resolved pixel in the output image.

If OCR is still unsuccessful post-SRCNN processing, the candidate region determination is repeated after inverting the colours of the original image.

Identifying and Measuring the Scale Bar. The scale bar in the image is then identified via a two-step procedure. The first step involves a search through all candidate regions of the image that OCR found to include text and that satisfied the aforementioned imposed conditions of text composition. For example, text such as '1 μm' is characteristic of textual information held within a scale bar. The search for the actual scale bar begins by expanding the candidate region of the text-containing imagery evenly by 9, keeping the original candidate region in the centre (see Figure 3.6), save one exception. This exception concerns any region that was previously determined to be an artificial box, in which case the expansion is unnecessary, and this step is thus skipped.

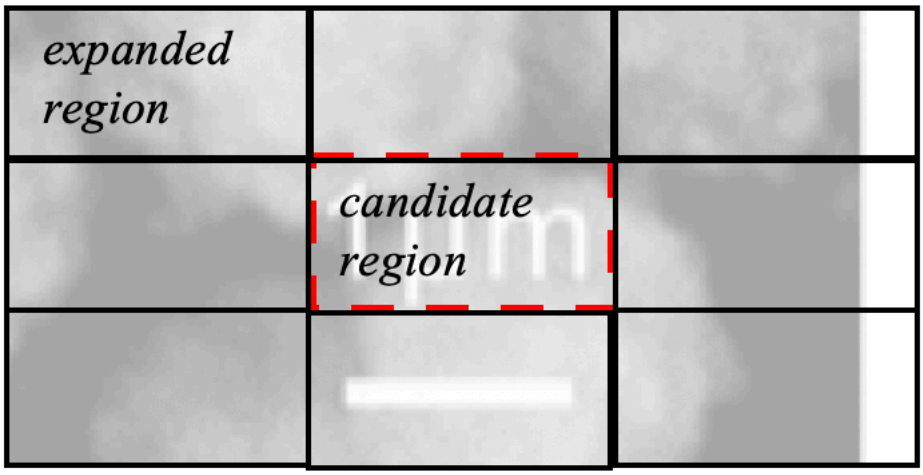


Figure 3.6 *Illustration of an expansion of the candidate region for a scale-bar search. A "ghost image" is embedded behind the boxed areas to illustrate the region in which the textual scale information will reside and provide an example of where the associated scale bar might be found in a neighbouring search-box region.*

The expanded candidate regions of text-containing imagery are then searched for objects which possess highly regular rectangular contours, i.e. befit scale bars. This high level of

regularity is imposed by stringent acceptance criteria in the scale bar search, whereby rectangular contours must exhibit (i) high aspect ratios (width / height) of at least 1.3, and a width less than the total image width / 1.5; (ii) either the mean pixel intensity of the contour must be greater than 200 or less than 30 (white or black); or the pixel intensity mode must be greater than 245 or less than 10. Note that it is important to analyse the colour of each candidate region via two colour-contrasting methods since the colour of a small scale bar can be easily affected by image resolution and contrast.

Although there is usually only one region of the image that contains the scale bar, sometimes multiple candidate regions of text-containing imagery can reach this stage of image processing; such instances typically involve images that contain additional text inlays or 'foreign' scale bars - for example, where a small image containing a scale bar is inset within the main image which contains the scale bar sought. The second procedural step of scale bar identification thus resolves these ambiguities. The case of additional text inlays is easily dealt with since such regions do not contain a scale bar. Thus, once can search the region immediately surrounding this text for an object that has an approximately rectangular shape with a high aspect ratio. If no such shape is found, the candidate region is removed from further consideration. In the case where 'foreign' scale bars are present, all detected scale bars are measured and the largest is used since that is the one most likely to be associated with the main image.

Calculating the Scale of the Image. The {text,contour} pair describing the scale of the image have now been identified. The width of the contour describing the scale bar (in pixels) together with the textual value and unit are then used to calculate the scale of the actual image. This is achieved by simply referencing the number of pixels along the scale bar to its annotated text and unit and then normalizing this reference to the number of pixels in the entire image. Note that our *ImageDataExtractor* code hardwires the text resolution of the unit in each scale bar to be either μm or nm, since this restriction engenders data extraction with high precision, yet negligible compromise on recall; optical microscopy images nearly always carry one of these two unit options. The scaled image output is presented in units of pixels/m in order to comply

with the *Système International* (SI) *d'unités* for length. If the scale of the image cannot be determined, detection continues with all outputs reported in units of pixels.

Figure 3.7 presents an example of an image whose scale information has been interpreted by *ImageDataExtractor*, using the overall procedure described in this section. This image will be familiar as it is 'image (c)' within the image panel shown in Figure 3.2. For the benefit of the reader, this image will be used as the running example throughout this chapter, to illustrate the function of each salient processing step of *ImageDataExtractor*.

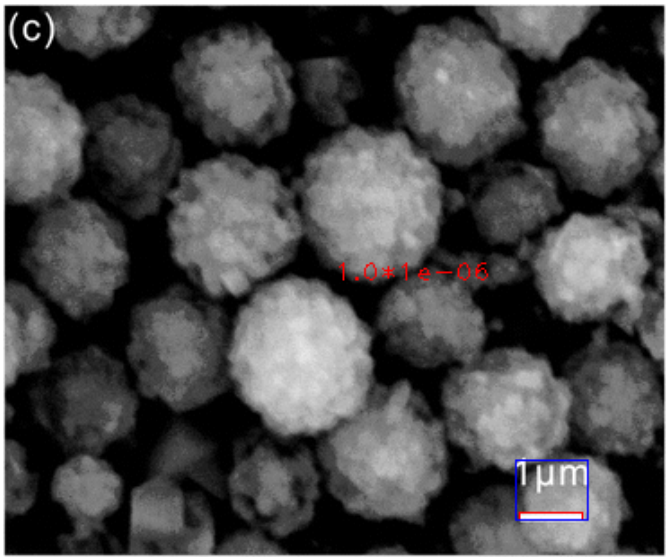


Figure 3.7 *Demonstration of the scale interpretation function in ImageDataExtractor, using the running example image [108]. The output of ImageDataExtractor provides image annotations in red and blue. These show that the value of the scale bar has been interpreted correctly, as shown by '1.0*1e-06', given in red; the scale bar has been correctly defined by the red rectangle; and the region containing scale information region is demarcated concisely by the blue rectangle.*

Particle Detection. Having interpreted the scale information, *ImageDataExtractor* then undergoes a process that detects individual particles within an image. Figure 3.8 shows how this process breaks down into its constituent parts. The process begins with a thresholding procedure that samples each pixel in an image to determine if it belongs to particle information or not.

Pixels belonging to particles are coloured white while those pertaining to the image background are coloured black; the resulting image binarisation highlights the visual contrast of particles, to facilitate their detection which comprises the next step. Particles are provisionally detected using the same border detection algorithm by Suzuki *et al.* [110] that was described earlier for scale interpretation. Particles are then approximated to polygons using an algorithm, whereupon a range of corrections is employed to assess if these detected particles are: (i) real or not (false positive detection), (ii) single particles or particle clusters which need to be broken apart, or (iii) particles that have been truncated (e.g. by the image border, or by image contrast problems). The list of detected particles in an image is confirmed once *ImageDataExtractor* has processed this range of corrections, whereupon they are taken on to quantitative analysis (see next section). All image manipulations were carried out using built-in OpenCV[109] functions or manual array editing, unless stated otherwise. Each of these particle detection and corrections steps are now described in detail.

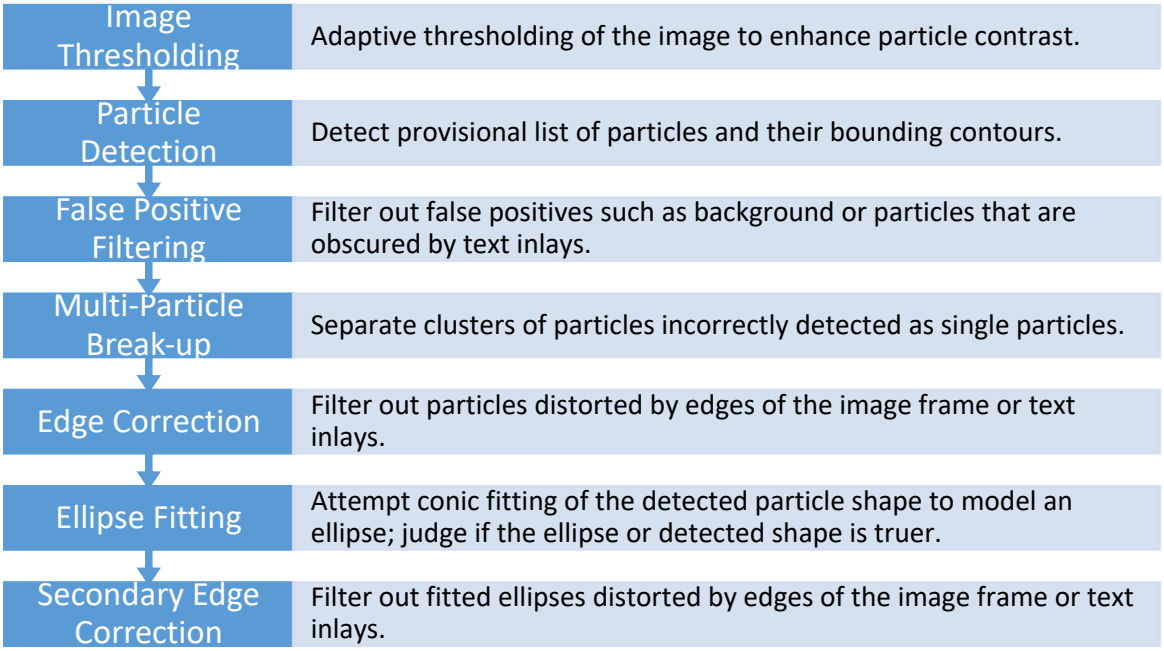


Figure 3.8 *Process for detecting individual particles in images.*

Thresholding. This procedure generates a black and white version of the original image by assigning particle information as foreground (white) against background information (black). This colour classification is achieved by sampling each pixel of the image with an adaptive thresholding procedure which considers the absolute value of its colour encoding, but relative to the colour values of its neighbouring pixels. Nearby pixels with similar colour values are likely to belong to the same classification (e.g. particle or background), while a particle boundary is likely to lie between neighbouring pixels with disparate colour values. Subject to these local similarity considerations, the absolute value of the colour encoding is compared against a threshold value to determine whether a given pixel is a particle (white) or background (black).

The precise algorithm used for this adaptive thresholding procedure is built into OpenCV via the function, adaptiveThreshold, which is used with two parameters: THRESH_BINARY, which sets pixels brighter than the threshold value to the maxValue, which in this case is white; and ADAPTIVE_THRESH_GAUSSIAN_C, which calculates the mean of the neighbourhood window using a Gaussian kernel, as opposed to the default statistical mean. The thresholding can be described by the equation below:

$$p_f = \begin{cases} 255, & p_0 > T \\ 0, & p_0 < T \end{cases} \tag{3.3}$$

where the threshold value, T, is a weighted sum (cross-correlation with a Gaussian kernel) of the 151 x 151 pixel neighbourhood of the pixel coordinates.[109]

This type of thresholding is referred to as adaptive because it considers its neighbouring environment, in contrast to, say, absolute thresholding which would simply label pixels as black or white according to whether the pixel colour value was above or below a fixed threshold value. Locally adaptive thresholding methods, such as those in 'category 6' as defined by Sezgin *et al.*[119] are good for classifying 3-D particles rendered in 2-D images, since their 3-D nature imparts different "lighting" conditions across different local regions within an image. It also

ensures that particles which are darker, because they lie deeper in the surface or they possess a disparate composition to other particles, are not classified as part of the background.

Following adaptive thresholding, noise reduction is applied using a technique called median filtering, which sets all pixel intensities in a window (which we have set to 3 x 3 pixels) to the median of that window. This filtering process uses the medianFilter built-in function of OpenCV.[109] There are various reasons to employ noise reduction routines. As can be surmised from our small window size, our motivation to employ this step is to reduce "salt and pepper" noise in the background rather than to perform its more common functions such as smoothing out contour features.

Figure 3.9 presents the result of thresholding the running example image in this chapter, using *ImageDataExtractor*.

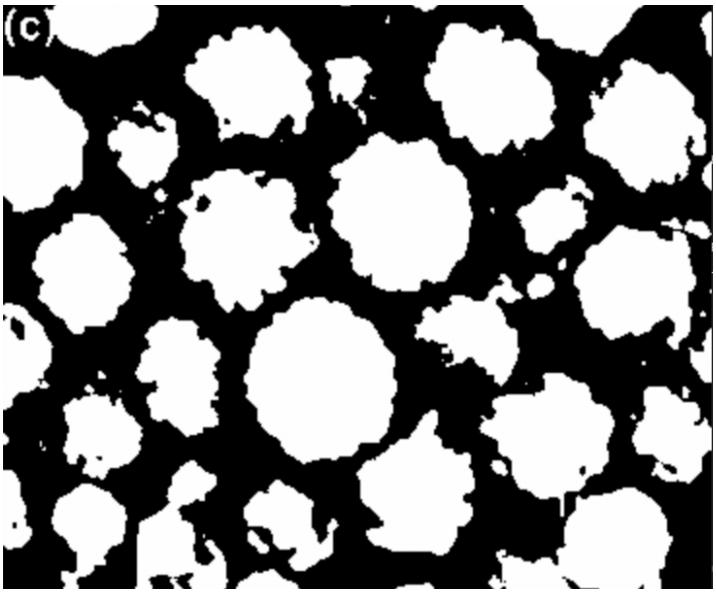


Figure 3.9 *The running example image [108] after adaptive thresholding and noise reduction. Please note the figure label (c) was carried over from the original publication and can be disregarded for our purposes.*

Detection. Individual particles are now defined according to image regions that are contained within a closed contour. Such contours were identified using the same contour detection functions that were outlined for the aforementioned scale interpretation process. The number of points that define each contour in the raw output of this detection procedure can be large. Fewer points can often describe contours determined by digitalised methods, without significant loss of information. To this end, each contour was approximated to a polygon, using the Douglas-Peucker algorithm.[120] Each contour was fitted to within 0.25% of its perimeter, to optimise efficiency.

The result of applying this particle detection feature in *ImageDataExtractor* to the running example image in this chapter is illustrated in Figure 3.10.

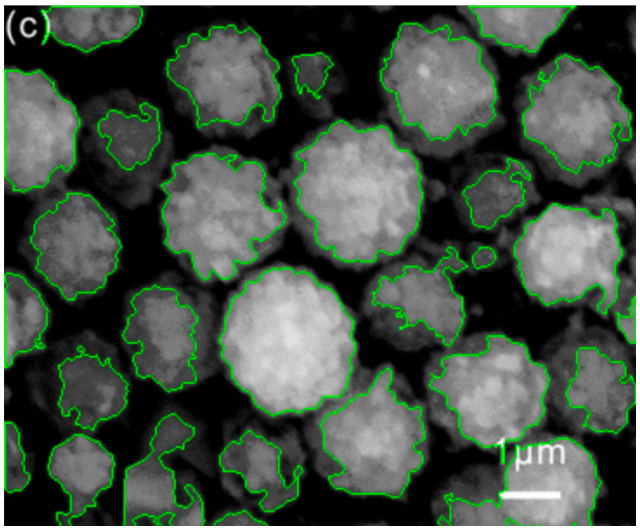


Figure 3.10 *The profiles of all closed contours (green) detected in the running example image [108] by the contour detection procedure within ImageDataExtractor, prior to the subsequent correction steps in the detection process Please note the figure label (c) was carried over from the original publication and can be disregarded for our purposes..*

Following this procedure, almost all of the contours required to define each particle in the image have been identified. However, certain types of diagnostics need to be performed at this stage in order to pick out contours that are either deformed, or present as false positives. The

display of overlapping particles in an image, or poor image compression, are typical causes of such issues. In fact, the purpose of the remaining steps of the particle detection process (as shown in Figure 3.8) is to filter out and, if necessary, modify the preliminary listing of particles from the contour detection procedure to ensure correct and optimal identification of every particle.

False Positive Filtering. The first correction step involves a basic identification, and thence elimination, of contours that are false positives, *i.e.* contours that appear to describe particles but are in fact artefacts. Such false positives are often caused by artificial inlays (text, shapes, arrows etc.) or sometimes regions of textured background. The correction procedure involves first calculating the mean, mode, and standard deviation of the grayscale pixel intensities of each pixel contained within the borders of a detected contour. Contours are then eliminated wherever their pixels meet the following statistical criteria: the mode pixel intensity is greater than 250, or the mean pixel intensity is less than 2, greater than 240, or less than 60% of the global mean, and the standard deviation in pixel intensity is smaller than 25% of the global value.

The running example image in this chapter did not actually contain any false positives, so somewhat anomalously, a different example [121] is shown in Figure 3.11 to demonstrate the effectiveness of this particular step using *ImageDataExtractor*.

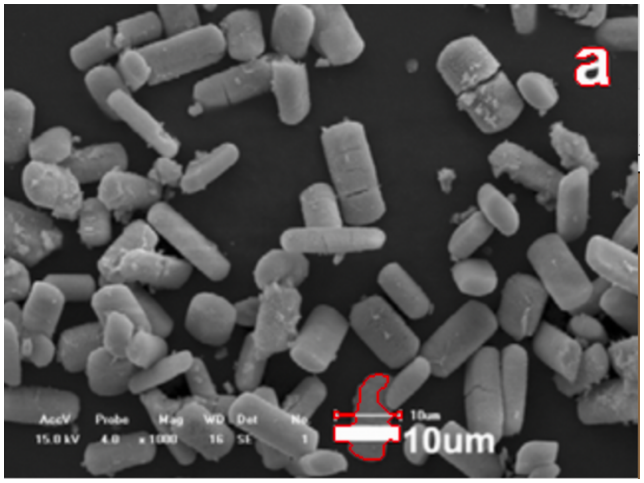


Figure 3.11 *Determination of false positive contours in ImageDataExtractor, demonstrated via an example image, reproduced with permission from reference* [121] *. Contours shown in red refer to two text inlays that cause these false positive contours: (i) the 'a' annotated on the top right corner of the image; (ii) the scale bar information at the bottom of the image. The false positive correction step in ImageDataExtractor eliminates both of these contours. However, it is worth mentioning that the false positive contour associated with the scale bar would have also been eliminated in a subsequent edge correction step (vide infra) that lies within the overarching detection workflow of ImageDataExtractor. Please note the figure label (a) was carried over from the original publication and can be disregarded for our purposes.*

Breaking up clusters. This step aims to separate particles that overlap or lie extremely close together in an image. The particles involved cannot be defined by closed contours as their borders are ambiguous; rather, the contour detection procedure erroneously categorises multiple particles as one large closed contour, i.e. as a single large particle. These clusters of particles are separated by a three-part sequence of operations.

Firstly, the multiple particles that have been erroneously classified as one particle are identified. This procedure uses the fact that the size of a contour for such false positive single particles will be anomalously large, relative to a true positive single particle, since they really comprise multiple particles. Accordingly, all contours that are larger than 50% of the mean contour area, or are 50% of the image size, are classified as possible multiple particles.

Secondly, features called ‘narrow bottlenecks’ are sought within these identified multiple-particle regions. Narrow bottlenecks are characteristic features of an object within an image that connect particles by a thin channel or ‘neck’. The points associated with the smallest diameter of this neck are called convex defect points, which lie a significant distance from the convex hull of that object. This is perhaps easier to comprehend via an example. A narrow bottleneck within an object is illustrated via the insets of Figure 3.12. The accompanying geometric drawing of Figure 3.13 labels this bottleneck and the convex hull that surrounds the object; a convex hull is the polygon that wholly encapsulates an object while occupying the minimum area, pending the constraint that the contour of the object touches all vertices of the convex hull, whose reflex interior angles are all less than 180°. A narrow bottleneck is identified by first finding the convex hull of the object, using an algorithm described by Sklansky [122] which employs the convexHull function of OpenCV. If a narrow bottleneck exists within an object, then the shortest diameter of its neck will be bounded by two convex defect points. Such a point is formally defined as the point on the contour of the object which lies furthest in perpendicular distance from the convex hull boundary between two adjacent convex hull vertices. This is illustrated for a particle cluster in the running example image,[108] whereby Figure 3.13 denotes the furthest normal distances between the object contour and the convex hull, d_1 and d_2, between the convex hull vertices, v_1 and v_2, and v_5 and v_6, respectively. All convex defect points of an object can thus be calculated via standard geometric functions that determine the maximum distance between the object contour and the convex hull, between adjacent convex hull vertices. The close proximity of two convex defect points within an object, and their relative geometric positioning at opposite sides of the object, are considerations used to identify a narrow bottleneck.

If such points are detected and thus a narrow bottleneck is located, the particle cluster is sliced apart at these points and reinserted into the master list of identified particles in their broken-apart forms. This slicing operation represents the third and final part of the particle cluster separation algorithm. Figure 3.12 shows how *ImageDataExtractor* follows this process of particle separation for a particle cluster, using the running example[108] for the purposes of

demonstration. As an addendum, it is worth noting that samples sometimes include particles of multiple shapes that are actually physically joined; however, these sorts of clusters rarely feature narrow bottlenecks, and thus survive this step unharmed.

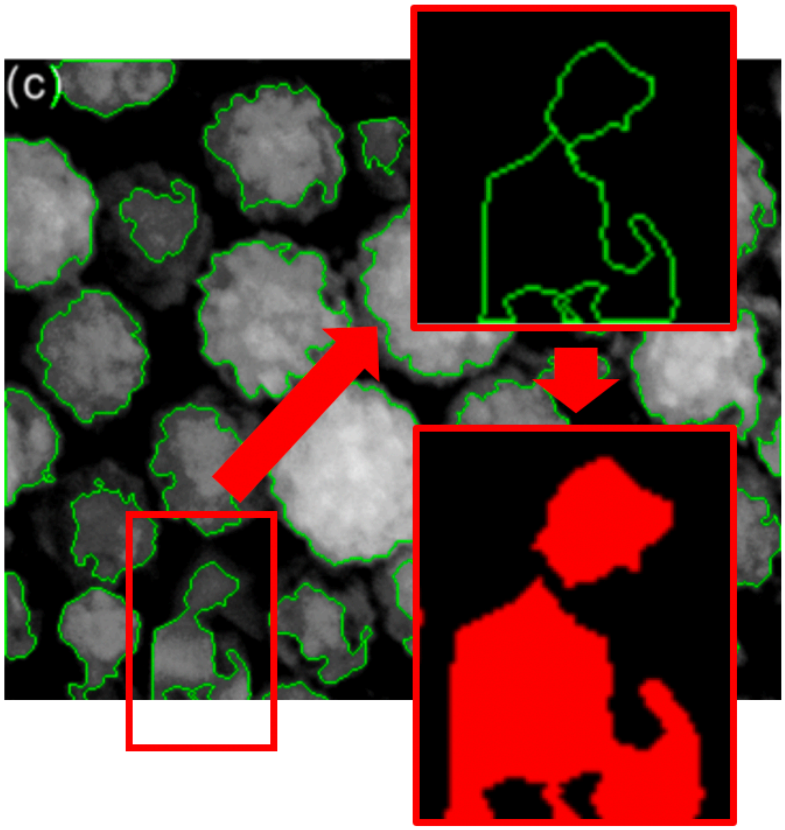


Figure 3.12 *Demonstration of the step in ImageDataExtractor that breaks up particle clusters using the running example* [108] *carried throughout this chapter. This process is applied to almost every particle. The insets with red borders show the sequence by which the particle cluster is identified and broken up. Please note the figure label (c) was carried over from the original publication and can be disregarded for our purposes.*

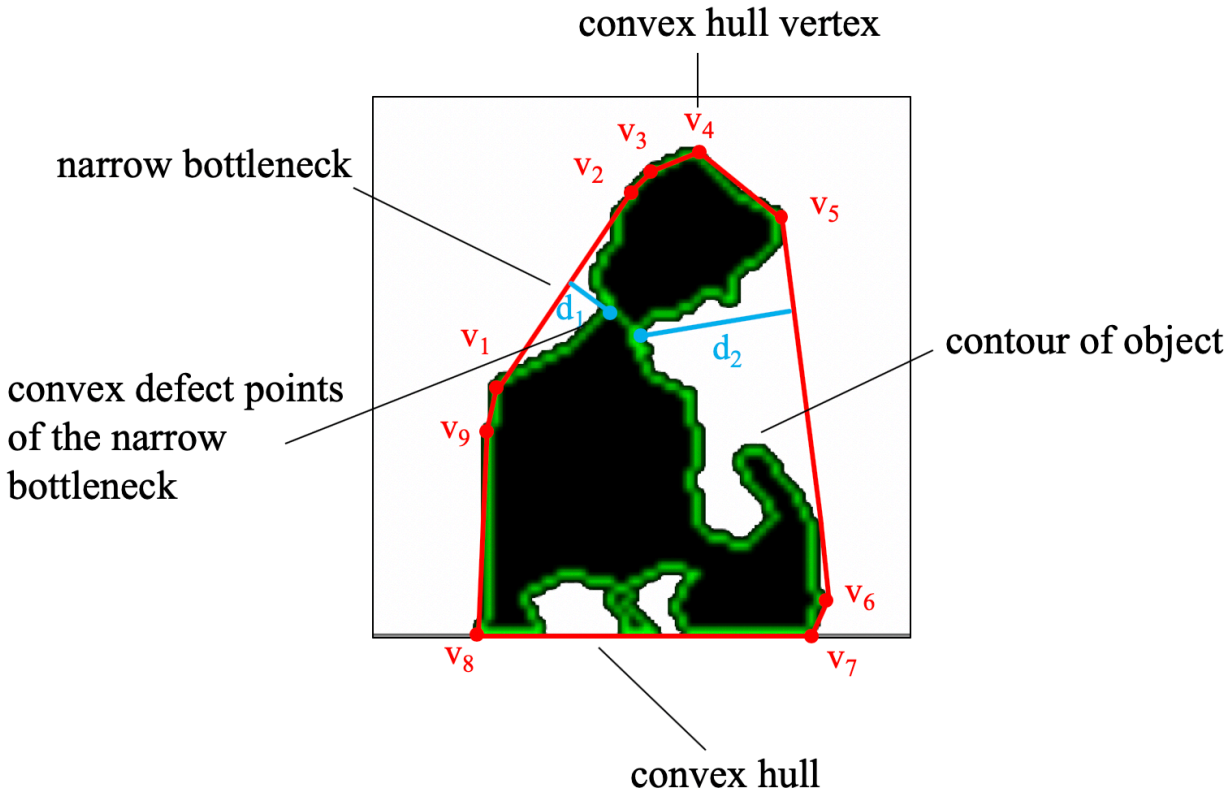


Figure 3.13 *This schematic drawing shows the geometry associated with the cluster break up algorithm. Drawing annotations include the convex hull (red), convex hull vertices, v_n, (red) convex defect points of the narrow bottleneck (blue), and the associated normal distances, d_1 and d_2, between the object contour (green) and the convex hull between adjacent convex hull vertices, v_1 and v_2, and v_5 and v_6, respectively.*

Edge Correction. The edge correction step eliminates any particles deemed to be artificially distorted by the borders of the image frame or scale information region. It is a simple process where any contour featuring two or more vertices intersecting the image borders or scale bar region are eliminated from the master list of identified particles. The effect of performing this correction in *ImageDataExtractor* is illustrated in Figure 3.14 using the running example[108] used throughout this chapter.

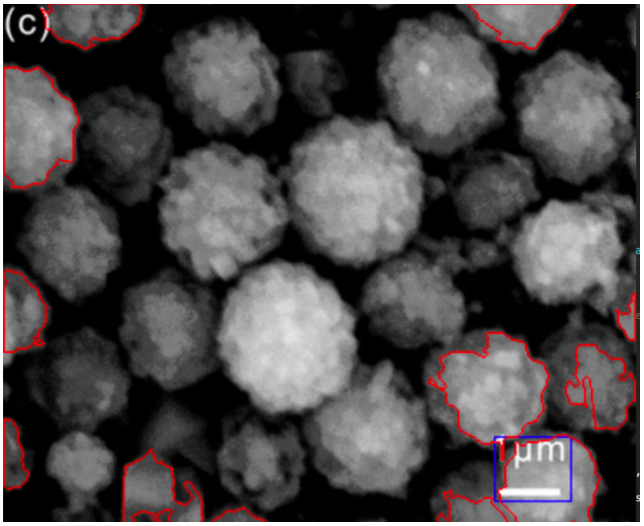


Figure 3.14 *Demonstration of the edge correction procedure in ImageDataExtractor, using the running example* [108] *used throughout this chapter. Particle contours that intersect the borders of the image or the scale information region (blue rectangle) are eliminated. Please note the figure label (c) was carried over from the original publication and can be disregarded for our purposes.*

Ellipse Fitting. This penultimate correction step in particle detection tackles the pervasive problem of non-uniformly shaded particles. In such cases, the darker portions of a particle may end up being ignored by the border contour while its lighter side is defined correctly. A hypothetical example is given in Figure 3.15, where the original particle (left) is detected poorly, as a shape resembling an ellipse truncated by a chord (middle); the chord bisects the registered (lighter) part of the particle from its curtailed darker portion. Conic fitting is used to model the shape profile of the partially detected particle to that of a Euclidean conic section, whereby this geometric approximation projects well the shape of the full particle (right). Thereby, the near-elliptical part of the originally detected particle in our hypothetical example is fitted as an elliptical conic section which envelopes the originally detected particle; the conic fitting method used in *ImageDataExtractor* applies the OpenCV conic fitting functions, which are derived from the procedure described in first part of the summary paper by Fitzgibbon and Fisher.[123]

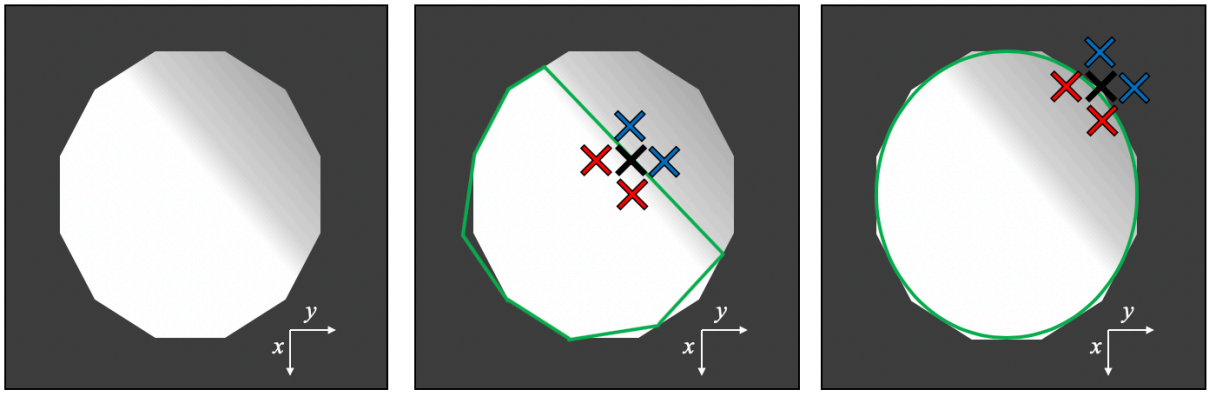


Figure 3.15 *(left) A heterogeneously shaded particle, (middle) the particle that has been erroneously detected (green contour), owing to detection issues from the heterogeneous shading of the particle, (right) conic fitting of the originally detected particle to an ellipse (green contour).*

The originally detected particle and the fitted elliptical conic section are compared via a custom metric that we define as the *discreteness index* (DI). This judges autonomously which option is the better fit to the true particle. The fitted elliptical conic section is approximated to a polygon with an angle of 175° in between subsequent line segments (using the OpenCV function ellipse2Poly with delta parameter set to 5) for the sake of efficiency, and is thus defined by the vertices of a polygon. The DI of a contour, DI_{cont}, is defined to be the mean pixel intensity difference between two pairs of points just inside and just outside the contour border at every vertex, whereby:

$$DI_{vertex}(x,y) = \begin{cases} I_{((x+3),y)} - I_{((x-3),y)}, & if\ x_i < x_m \\ I_{((x-3),y)} - I_{((x+3),y)}, & if\ x_i > x_m \end{cases} + \begin{cases} I_{(x,(y+3)} - I_{(x,(y-3))}, & if\ y_i < y_m \\ I_{(x,(y-3)} - I_{(x,(y+3))}, & if\ y_i > y_m \end{cases} \tag{3.4}$$

$$\mathrm{DI}_{cont} = \frac{\sum_{(x,y)_0}^{(x,y)n} DI_{vertex}(x,y)}{n} \tag{3.5}$$

for all vertices (x,y) that define the particle contour, where $I_{(x,y)}$ is the intensity of the pixel marked by the vertex with coordinates (x,y), (x_m,y_m) are the coordinates of the centre of mass of the contour and n is the number of points in the contour. In practice, the fitting process is applied to the convex hull of the contour instead of every vertex of the fully detected contour, to allow accurate results with reduced computing cost.

Using Figure 3.15 (middle) as an example, the value of interest (DI_{vertex}) for the vertex marked by the black cross (x,y) would be calculated as follows. Since this vertex (x,y) has $x < x_m$ and $y > y_m$ (the point is above and to the right of the centre of mass of the contour), DI_{vertex} would be calculated by subtracting pixel intensities of the points marked by the blue crosses (above and to the right) from that of those marked by the red crosses (below and to the left).

$$DI_{vertex}(x,y) = (I_{((x+3),y)} - I_{((x-3),y)}) + (I_{(x,(y-3)} - I_{(x,(y+3))}) \tag{3.6}$$

i.e. (red_x – $blue_x$) + (red_y – $blue_y$)

The discreteness index is the mean of all such values for every vertex, i.e.

$$DI_{cont} = \overline{DI_{vertex}} \tag{3.7}$$

One can imagine this calculation repeated for multiple points along each contour in the hypothetical case presented in Figure 3.15, where the original contour insufficiently described the particle (Figure 3.15, middle). The elliptical conic section fitted to the near-elliptical portion of the original particle shape (Figure 3.15, right) would have a higher discreteness index than the

original fitted profile in Figure 3.15 (middle), since the difference in colour on either side of the contour is, on average, larger for the fitted ellipse. The elliptical conic section would thus be determined to represent the more correct fit to the shape of the true particle.

Considering now a real example, using the running image used throughout this chapter,[108] Figure 3.16 illustrates the effect of fitting ellipses to particles using *ImageDataExtractor*. This demonstrates that the fitting method is very powerful, despite its simplicity. While this simple method of determining pixel positions by moving by 3 pixels vertically or horizontally (to determine sample points inside/outside the contour) does not work perfectly in the case of oddly shaped particles, it is deemed sufficiently accurate, such that the computational cost of a more complex approach is not justified.

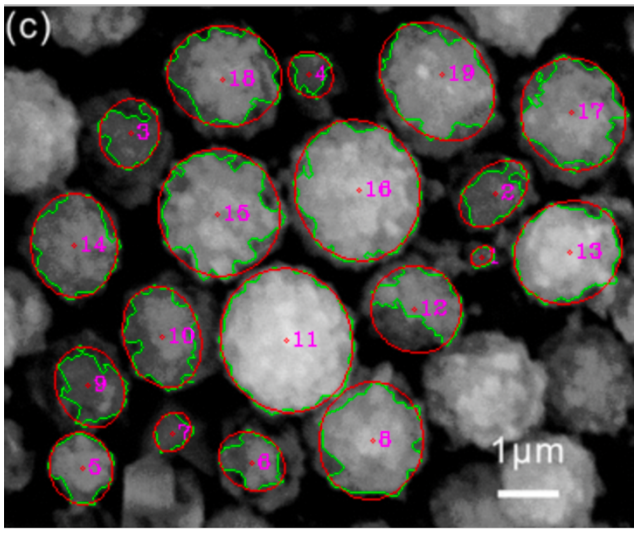


Figure 3.16 *The results of fitting ellipses (red) to the contours that described the originally detected particle shapes (green) in the running example image [108] used throughout this chapter, as implemented by ImageDataExtractor. The shape of the originally detected contour is compared with its fitted ellipse; the fit that best matches the true shape of the particle gets passed to the final product. The fitted ellipse is deemed to be the better fit in this image for particles 10, 11 and 14, whose labelling is given in magenta. Please note the figure label (c) was carried over from the original publication and can be disregarded for our purposes.*

Secondary Edge Correction. The final correction step is a secondary edge correction step, identical to the previous one, which serves to eliminate any new elliptical fits that intersect with the image frame or scale bar region. The updated master list of border contours defining each particle can now be passed on to the quantitative analysis stage of *ImageDataExtractor*, as described in the next section. Figure 3.17 shows the set of particles in the running example [108] used throughout this chapter that are finally selected for this next stage of analysis.

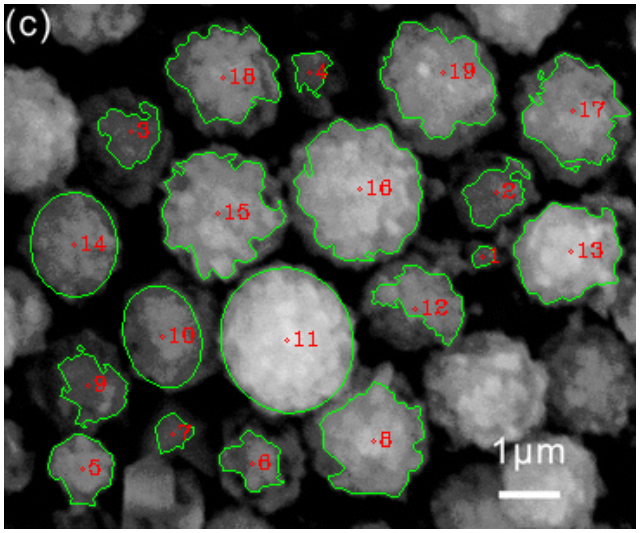


Figure 3.17 *Visual representation of the running example image [108] used throughout this chapter, annotated with the final set of detected particles that are taken forward for quantitative analysis in the next stage of ImageDataExtractor. Please note the figure label (c) was carried over from the original publication and can be disregarded for our purposes.*

Step 3: Quantitative analysis of material objects of interest in images including the creation of radial distribution functions

Microscopy images become much more useful in data analysis once representative particles contained within them have been detected, since such particles can be analysed quantitatively. *ImageDataExtractor* provisions for this utility by subjecting all detected particles from inputted images to a quantitative analysis process. This process comprises six steps which are summarised in Figure 3.18. The number of detected particles in each image processed by

Step 2 of the *ImageDataExtractor* system pipeline is first passed to this Step 3, as the source for quantitative analysis. These particles are then classified according to shape, by performing a similarity match on the contour of each particle with respect to a reference list of commonly recognised geometric shapes (e.g. circle, rectangle). A histogram that describes the range of particle sizes is then constructed. Particle whose sizes are statistical outliers are distinguished at this point, to afford a list of particles of representative sizes. This is to provide the user with an idea of the expected particle size, not necessarily for the individual sample, but for the material, filtering out unusually sized particles that are often incorrect detections or uncharacteristic of the material. However, further calculations, such as the mean aspect ratio of the particles, are applied to the entire list of detected particles, to more fully describe a given sample. Pending that at least 10 particles have been detected, a radial distribution function (RDF) of the particles in each image is then calculated in order to depict the separations between particles. This RDF is presented in the output of *ImageDataExtractor* together with the annotated image that visualises the particle detection results. The *ImageDataExtractor* output also contains a textual listing of the number of particles detected, the representative particle size, the shape resemblance of particles to common geometries, and the average aspect ratio of particles in each image. The figure and document from which each image is sourced is also identified in this textual output. A detailed description of the calculations that produce these output features is now given. A typical output of *ImageDataExtractor* is then shown for the running example image [108] used throughout this chapter.

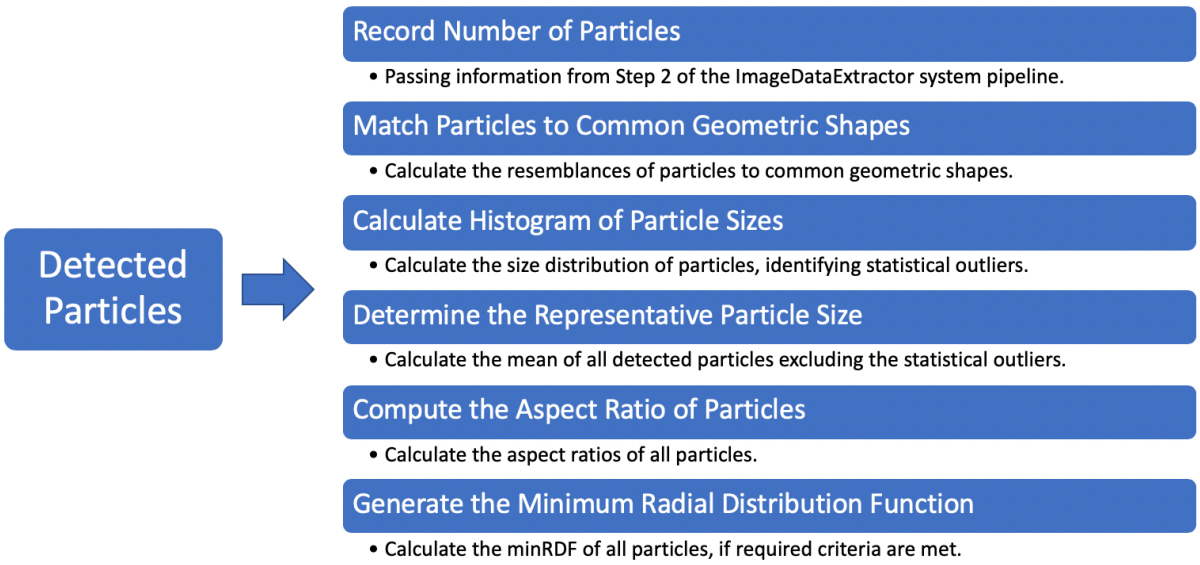


Figure 3.18 *Quantitative analysis steps in ImageDataExtractor.*

Match Particles to Common Geometric Shapes. A shape matching classification procedure is used to portray the resemblance of detected particles to regular geometric shapes. This is accomplished by comparing the shape of the detected contour for each particle with that of a reference list of common geometric shapes (e.g. rectangle, ellipse, diamond, circle), using the similarity index according to Hu moments.[124] Thereby, the similarity of two given shapes, I_1, is calculated according to:

$$I_1(A,B) = \sum_{i=1\ldots7} \left| \frac{1}{m_i^A} - \frac{1}{m_i^B} \right| \tag{3.8}$$

$$m_i^A = sign\left(h_i^A\right) \cdot \log\left(h_i^A\right) \tag{3.9}$$

$$m_i^B = sign(h_i^B) \cdot \log(h_i^B) \tag{3.10}$$

where m is the sign-adjusted logarithm for the i^{th} Hu moment; h_i^A and h_i^B being the Hu moments of A and B.[109] I_1 ranges from 0 to 1, where 0 denotes a perfect match. The Hu moments were calculated in *ImageDataExtractor* using the OpenCV function, matchShapes.

If the statistical mode of I_1, determined across all particles surveyed in an image, is less than 0.1 for any of the reference geometric shapes, the textual output of *ImageDataExtractor* notes the most common shape resemblance of the particles in the image. If multiple particle shapes manifest within an image, a range of shape resemblances of particles is also given in the textual output of *ImageDataExtractor*, together with their associated similarity indices. A default file of reference geometric shapes, shapes_to_match.png, is provided for *ImageDataExtractor*. However, the user can manually create their own reference shapes of interest, using the same .png image file format, and compare them against the particle contour shapes from input images.

Calculate Histogram of Particle Sizes. *ImageDataExtractor* generates a histogram of the particle size distribution for each image that it surveys. The radial unit for this distribution is given in either m^2 or px^2, since the originating image is a 2-D projection of a particle. m^2 is specified pending it has been possible to interpret the image scale; otherwise, px^2 is specified. This histogram is used to distinguish particles of representative sizes from those that are deemed to be statistical outliers. The removal of statistical outliers channels the range of detected particles to a representative particle size with modest standard deviation. This channelled metric is useful for experimental practitioners such as materials scientists who tend to be most interested in a typical size of nanoparticle for their work. However, it should be remembered that these are purely statistical outliers; particle sizes residing at the extreme values of a size range may not necessarily arise from incorrectly detected particles. Rather, the nanostructure of a material could naturally feature a range of particle sizes including many of a regular size and just a few that are much larger or smaller. Thus, while we provide the metric of representative particle size, the other quantitative analysis metrics presented herein are determined using all detected particles, irrespective of their size. Any size variations should naturally partition such metrics in any case; for example, particle separations that will be defined shortly will naturally separate large

particles from small ones; meanwhile, metrics that concern particle shape that have just been discussed are not size dependent.

ImageDataExtractor distinguishes a statistical outlier as any particle with contour area A, where:

$$A > \mu + 1.75 \times \sigma \text{ or } A < \mu - \sigma \tag{3.11}$$

with μ and σ being the mean and standard deviation of contour areas, respectively.

This process by which particles of representative sizes are distinguished from statistical outliers is illustrated in Figure 3.19 using the running example image [108] used throughout this chapter.

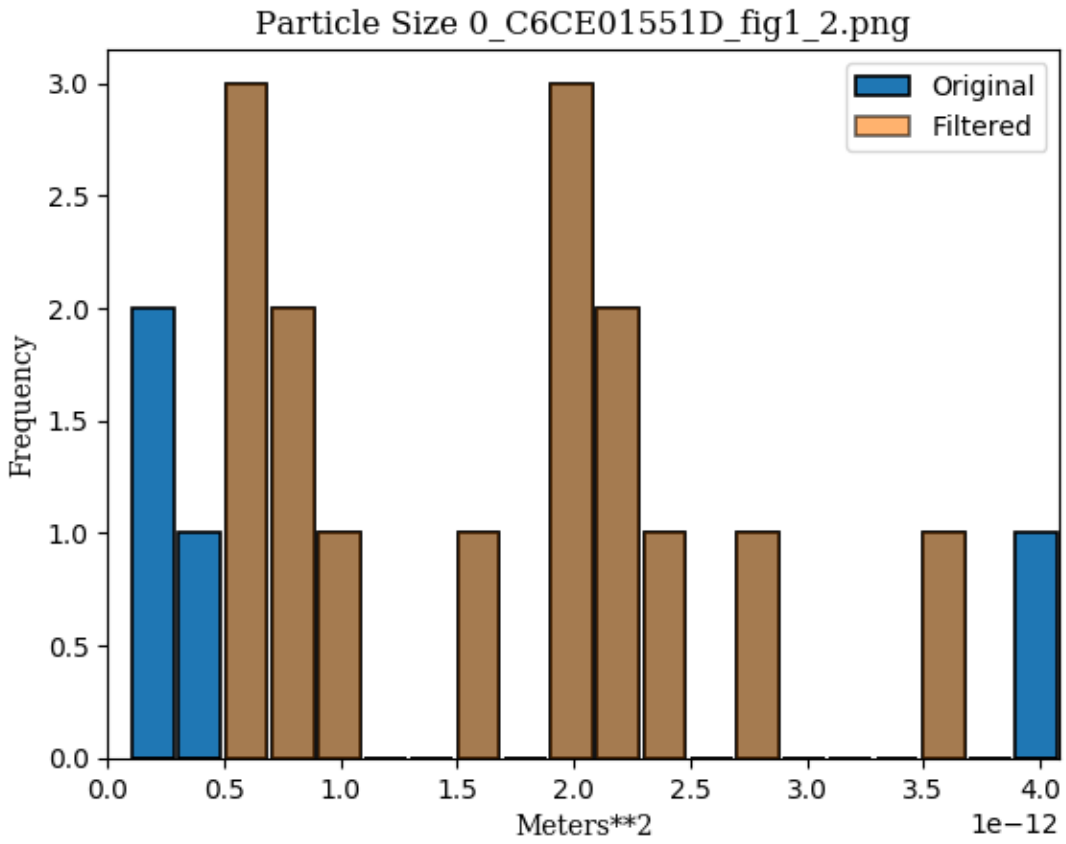


Figure 3.19 *Histogram of particle sizes (areas) for the running example image [108] used throughout this chapter. The brown columns are contributions from particles of representative sizes. The blue columns are contributions from particles deemed to be outliers.*

Determine the Representative Particle Size. This is simply a calculation of the mean of all particles of representative sizes in an image. Considering the filtered data in Figure 3.19, the representative particle size for the running example image[108] is calculated to be 1.92 x 10^{-12} m^2. In this case study, one can essentially approximate all particles to a sphere, such that their 2-D projected circular area translates to a mean particle diameter of 1.56 μm. This value is consistent with a judgement by manual inspection.

Compute the Aspect Ratio of Particles. The mean aspect ratio (width/height) was calculated using all particles from the image by calculating the bounding rectangle for each particle using the OpenCV function boundingRect and dividing the width of the rectangle by its height.

Generate the Minimum Radial Distribution Function. If at least 10 particles are identified in an image, then *ImageDataExtractor* computes a 'minimum radial distribution function' (minRDF) of particles which is also displayed as part of its output. In the context of this work, a radial distribution function (RDF) describes the average separation between the centroids of particles. The minRDF is defined herein as an RDF for particles detected in an image. This distinction is important since the omission of just a couple of particles from a conventional RDF can drastically alter it, while perfect particle detection in microscopy images cannot be guaranteed.

Thereby, the minRDF is calculated by counting the number of detected particles at a certain radial distance away from another, repeating for every particle. For each particle pair, the bounding rectangle of the secondary particle is identified via the presence of circles of various radii originating at the centre of mass of the primary particle. The radii of these circles range from: $[d - 1.5 \times r_2,\ d + 1.5 \times r_2]$, where d is the distance between the centres of mass of the particles and r_2 is the radius of the enclosing circle of the secondary particle. The default increment is set to 2 pixels. The radii of the circles which do intersect with the secondary particle are noted and added to a running list of intersections to radii. The process is repeated for each particle pair to generate the minRDF.

This process is depicted in Figure 3.20, using a particle pair from the running example image[108] used throughout this chapter. This visualisation also helps illustrate how, in addition to their distribution, an RDF can also provide insight into the shapes and sizes of particles.

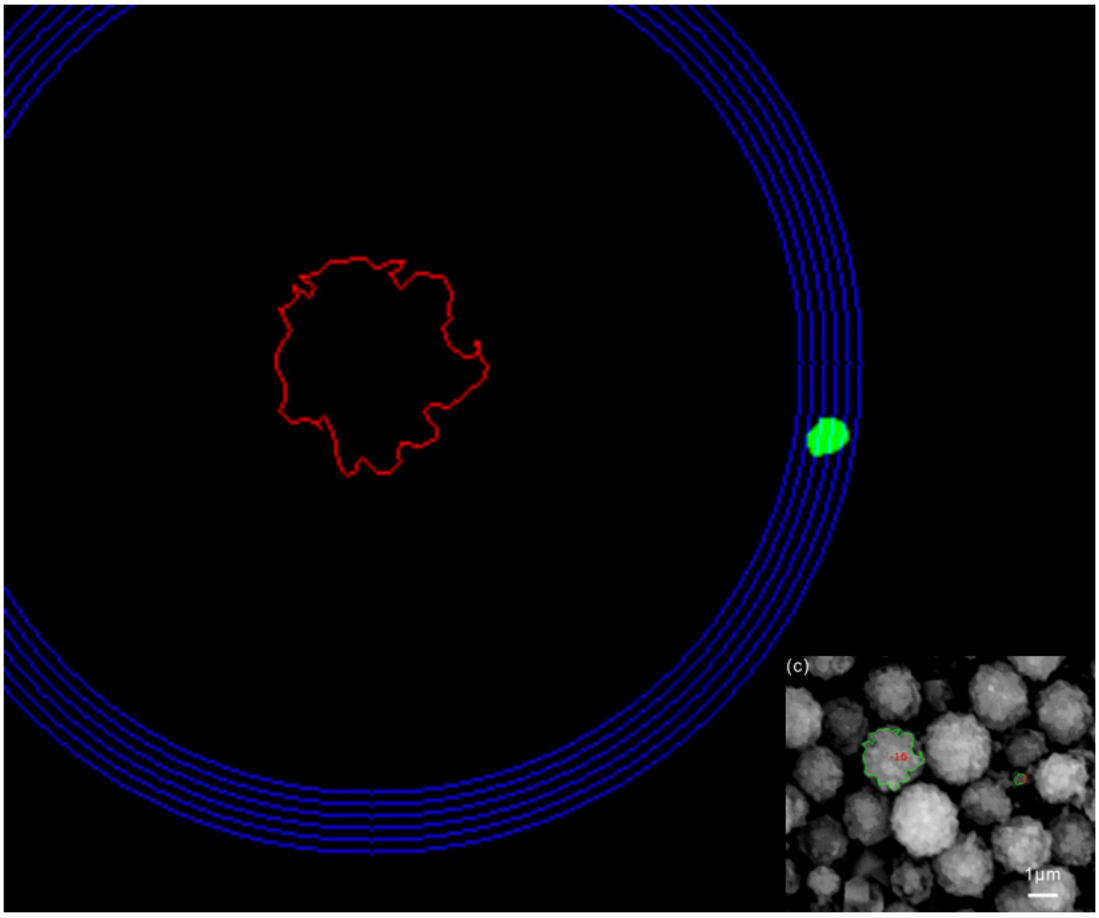


Figure 3.20 *Visualizing the determination of the bounding rectangle of the secondary particle in the minRDF generation process, using the particle pair 15 (red outline) and 1 (filled green) of the running example image [108] as labelled in Figure 3.16 (see inset) for the purposes of illustration. The radial distances analysed are shown as blue circles.*

Figure 3.21 presents the minRDF that is generated by *ImageDataExtractor* for the running example image[108] used throughout this chapter. Thus, the separation between particles in this case study ranges from 0.6 to 10.77 µm. The most common particle separation is 4.37 µm, while peak separations appear at 1.86, 2.89, 4.37, 5.06 and 6.09 µm.

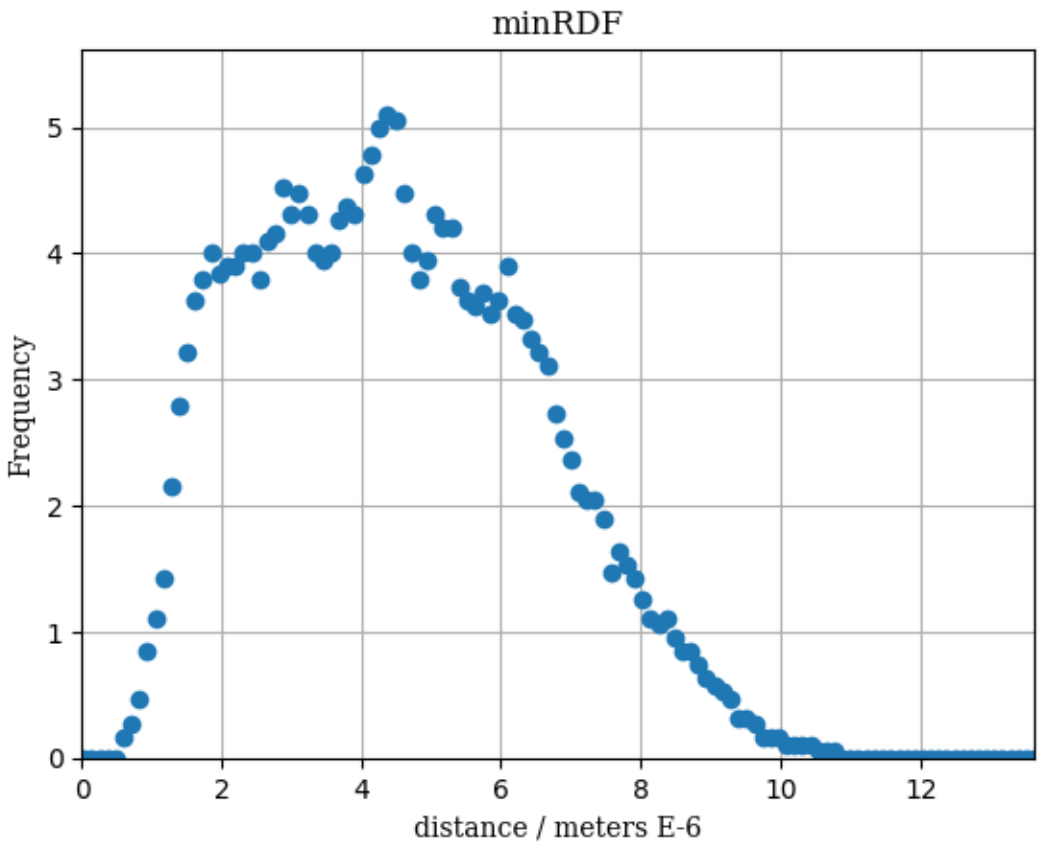


Figure 3.21 *The minRDF generated by ImageDataExtractor for the running example image* [108] *used throughout this chapter.*

Output Files from ImageDataExtractor. The minRDF is displayed in the *ImageDataExtractor* output, alongside its originating image that has been annotated with the particle detection results from Step 2 of its system pipeline. *ImageDataExtractor* also outputs the histogram of particle sizes for each image, the data of the minRDF in the form of a two-column text file and a textual summary of salient results from the quantitative analysis of each image. The corresponding textual output associated with this running example image is given in Figure 3.22.

```
0_C6CE01551D_fig1_2.png processed using ImageDataExtractor on
2019-05-07 17:36
Article DOI:
C6CE01551D
Figure number:
fig1

19 particles detected.
Representative particle size: 1.92e-12 sqm

Particle resemblances to regular shapes:
[('circle', '0.03'), ('ellipse', '0.89'), ('diamond', '0.87'), ('rectangle', '0.24')]
The 2D projections of the particles in this image most closely match a: circle
Average aspect ratio: 0.96
```

Figure 3.22 *Textual output of ImageDataExtractor for the running example image*[108] *used throughout this chapter.*

3.4 Technical Evaluation

The evaluation metrics, precision and recall, were calculated for four key stages of the *ImageDataExtractor* tool, according to:

$$precision = \frac{TP}{TP+FP} \tag{3.12}$$

$$recall = \frac{TP}{TP+FN} \tag{3.13}$$

where TP (true positives) are defined as the number of accurate outputs, FP (false positives) are the number of inaccurate outputs and FN (false negatives) are the number of null outputs which should not have been null.

Four stages of the *ImageDataExtractor* system pipeline were assessed by this evaluation. These were naturally partitioned according to its aforementioned distinct steps of image

processing: the figure splitting (Step 1), scale identification (the first distinct sequence of operations in Step 2), particle detection (the initial part of the second distinct sequence of operations in Step 2), and individual particle accuracy (the remaining sequence of operations in Step 2). These terms were defined earlier except for the individual particle accuracy; this is a measure of the success of detection on the individual particle level and describes how close the detected borders of the particle are to the true borders. It is calculated by manually counting the number of pixels included within the borders which are either correct or incorrect and any pixels that are falsely excluded.

The test set used for the evaluation consisted of 298 images that were sourced from papers in the academic literature and which all present therein as figure panels. *ChemDataExtractor*[13] was used to extract these images from a corpus of scientific articles comprising 6807 documents from the Royal Society of Chemistry (RSC) publisher. The extraction process employed a search query across all RSC journals for the words 'SEM', 'TiO2' and 'nano'. It should be noted that this query was meant to merely guide the collection process towards imaging-related scientific articles and did not limit the scope of imaging techniques or sample materials in question, since the criteria were not enforced on individual figure captions. The query was performed using RSC's 'Advanced Search' API which selects articles most appropriate to the query using its own relevance metric. These articles were then scanned for microscopy images using the parser provided in *ChemDataExtractor* for *ImageDataExtractor* and the first 3000 figures were downloaded – this was then filtered down to 298 individual images. The filter comprised a manual pre-processing stage that imposed three acceptance criteria whereby at least one image in the figure must: (i) include at least one discernible discrete particle fully included within the frame of the image; (ii) include a legible scale bar, scale value and unit; (iii) not be obscured by large overlays or image insets. These image content criteria, as well as an image resolution floor of 50000 square pixels, were enforced as required throughout the evaluation steps. DOIs for the 298 filtered images and their original Figure numbers, within the documents from which they were obtained, can be found at: https://pubs.acs.org/doi/suppl/10.1021/acs.jcim.9b00734/suppl_file/ci9b00734_si_001.xlsx

Table 3.1 *Evaluation metrics for key steps of the ImageDataExtractor system pipeline.*

Evaluation Step	***Precision***	***Recall***
figure splitting	73.9%	80.2%
scale identification	81.1%	72.3%
particle detection	82.9%	61.1%
individual particle accuracy	99.5%	80.5%

Table 3.1 summarises the overall results of this evaluation for each of the four assessed steps of the *ImageDataExtractor* system pipeline. The precision is over 80% for all processing steps of individual images, while the precision of the figure splitting algorithm (73.9%) that generates these images for *ImageDataExtractor* processing is almost as good. The precision is generally considered to be the most important evaluation metric in this field of study, since the literature is so vast that there is safety in numbers: a high precision means that what is accepted is highly likely to be correct, and there are so many images available in the literature that one can deal with some shortfall in information where images are not assessed. Nonetheless, the recall which refers to the latter issue is still above 80% in two of the stages and decent (72.3%) for scale identification. The slightly lower recall for the particle detection stage is expected as this is, by far, the trickiest image processing step to perform, especially with regards to the segmentation and fitting steps. Overall, the results are very encouraging as the majority of the evaluation metrics exceed those of human error which is generally considered to be about 80%; while human labour is manual and tedious, *ImageDataExtractor* is a tool that has been tailored specifically to suit high-throughput image extraction and quantitative analysis tasks for microscopy. A more detailed breakdown of the evaluation, showing calculations of each metric for each figure in the evaluation test set, as well as source material, can be found at the link provided above.

3.5 Examples of Application

The running example image [108] used throughout this study contained near-spherical particles. However, it is important to demonstrate that *ImageDataExtractor* can work equally well on the processing of images containing a range of particle sizes and shapes. Figures 3.23-3.26 present the image input and *ImageDataExtractor* output for four further case studies to showcase the wider application of this software tool. Note that the original image in each case formed part of a collage of images but the full image panel is not shown for purposes of brevity. The original image panels are of course available in the cited article in the literature.

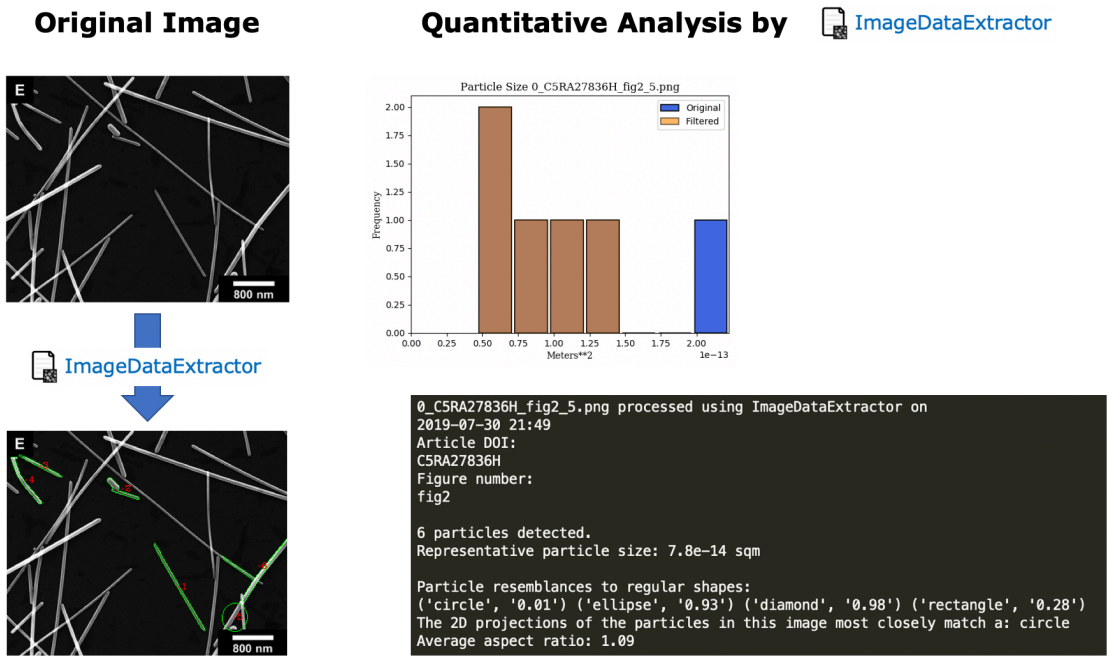


Figure 3.23 *(top left) The original SEM image of silver nanorods captured from panel 'E' of the original figure that comprised a 2 x 3 panel of images.* [125] *(bottom left) the particles detected by ImageDataExtractor in green, with their corresponding labels in red. (top right) Histogram of representative particle sizes. (bottom right) Textual output of ImageDataExtractor showing data source information and quantitative particle analysis. Note that no minRDF was calculated in this case since less than 10 particles were detected.*

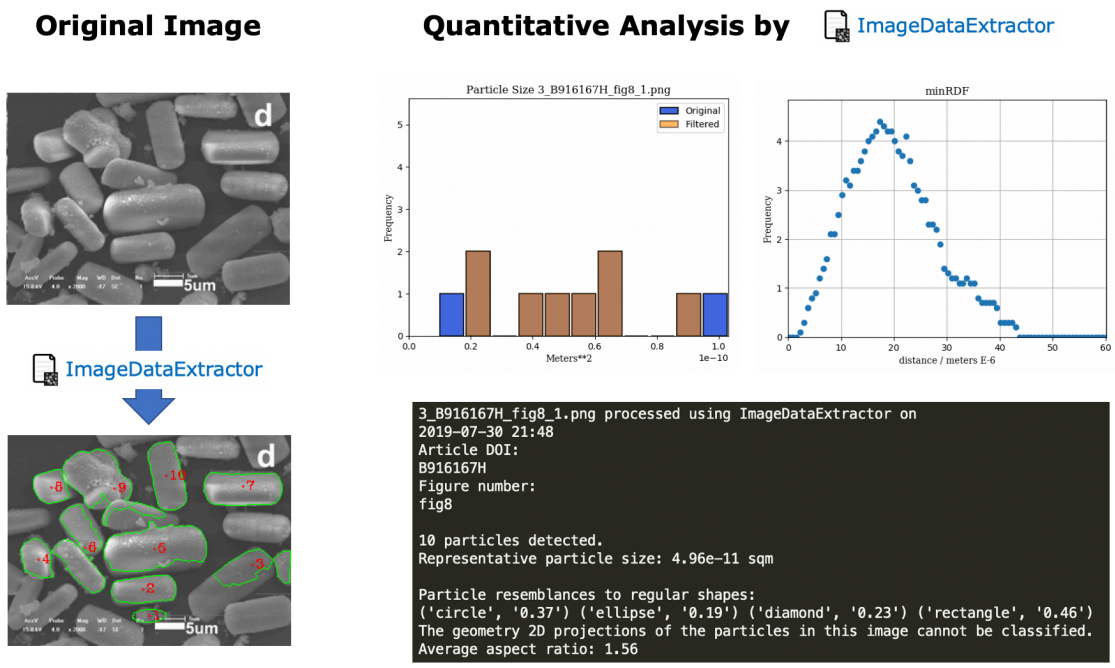


Figure 3.24 *(top left) The original SEM image of Ce-based precursors of asparagine:Ce^{3+}:oxalate, captured from panel 'd' of the original figure that comprised a 2 x 3 panel of images [121] (bottom left) the particles detected by ImageDataExtractor in green, with their corresponding labels in red. (top right) Histogram and minRDF of representative particle sizes and particle separations, respectively. (bottom right) Textual output of ImageDataExtractor showing data source information and quantitative particle analysis.*

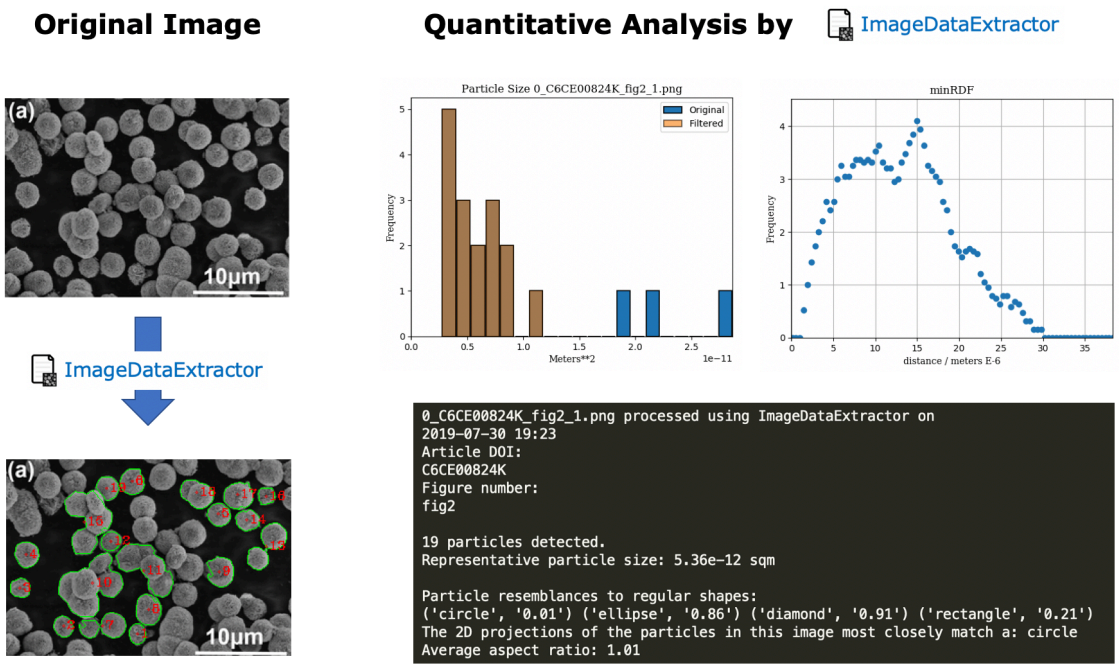


Figure 3.25 *(top left) The original SEM image of BiOBr, captured from panel '(a)' of the original figure that comprised a 2 x 2 panel of images.*[126] *(bottom left) the particles detected by ImageDataExtractor in green, with their corresponding labels in red. (top right) Histogram and minRDF of representative particle sizes and particle separations, respectively. (bottom right) Textual output of ImageDataExtractor showing data source information and quantitative particle analysis.*

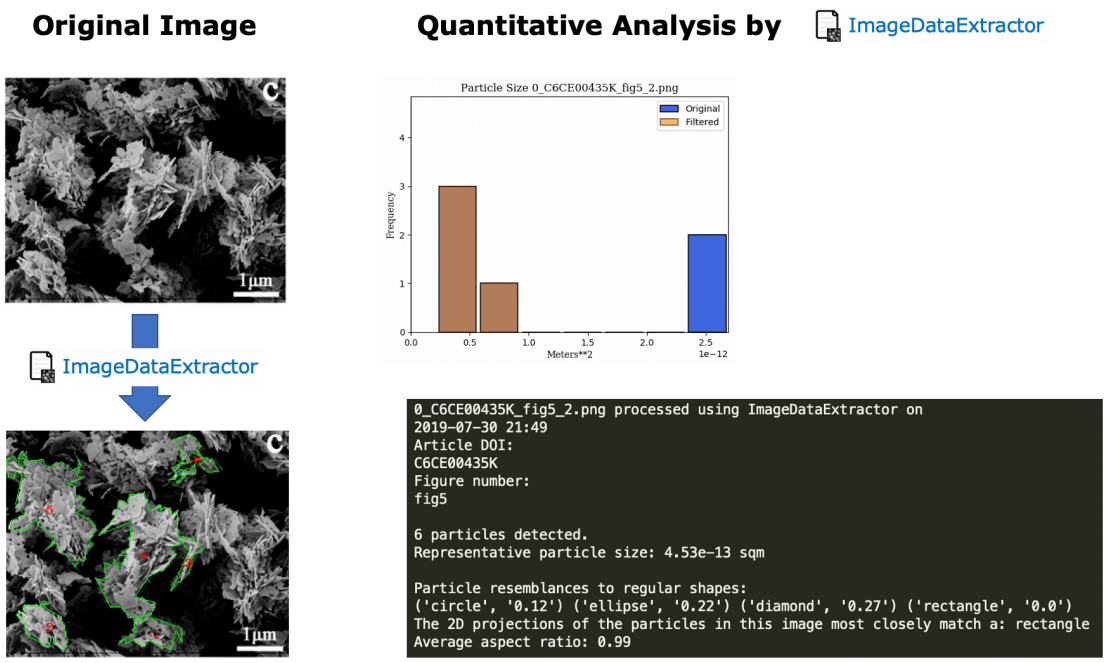


Figure 3.26 *(top left) The original SEM image of Bi-based carbonate clusters made up of nanosheets with a thickness of 5–15 nm obtained at 60 °C, captured from panel 'c' of the original figure that comprised a 2 x 3 panel of images.* [127] *(bottom left) the particles detected by ImageDataExtractor in green, with their corresponding labels in red. (top right) Histogram of representative particle sizes. (bottom right) Textual output of ImageDataExtractor showing data source information and quantitative particle analysis. Note that no minRDF was calculated in this case since less than 10 particles were detected.*

3.6 Conclusions

ImageDataExtractor is a software tool that auto-extracts microscopy images and analyses their content in a fashion that affords quantitative information about the shape, size and distribution of the 2-D projections of particles held within the image. The operational pipeline of the tool comprises three distinct steps: (Step 1) auto-identification and auto-extraction of microscopy images from documents whereby *ImageDataExtractor* calls upon its complementary text-mining tool, *ChemDataExtractor,* [13] for its data-mining capabilities; alternatively, it can act as a stand-alone tool on individual images. For cases where microscopy images are held within a collage of images within the figure of a document, this step also involves a separation process to

split the microscopy image from a panel of images. (Step 2) Decoding the scale information within a microscopy image using optical character recognition with help from super-resolution convolutional neural networks where required; and detecting the individual particles held in the image using various thresholding, segmentation, polygon fitting, and edge correction routines. (Step 3) Quantitative analysis of the detected particles within an image, afford statistical information about particle shape, size and separation. In contrast to other image-recognition based tools designed for microscopy images, *ImageDataExtractor* is capable of operating autonomously, in a high-throughput mode, and on a wide range of images. The tool and cognate code are also open source. Evaluation metrics calculated herein have shown that the tool performs well at all stages of image processing, and better than human error associated with manual image processing for the majority of its steps.

The high-throughput image extraction and quantitative analysis options of *ImageDataExtractor* mean that this tool is inherently poised toward our ultimate goal of auto-generating material databases of information on particles. Our applications for microscopy image information are naturally geared towards the expansive field of nanotechnology, given the particle sizes involved. Naturally, the tool has some limitations in its first incarnation; for example, it struggles with processing images that exhibit heavy depth cueing or highly heterogeneous contents and it only processes images that contain particles rather than bespoke objects (such as the microscopy imprint of an insect). Indeed, improvements in depth cueing and the ability to access heterogeneous content would improve the overall precision and recall of the tool. However, the generic design of our tool and the open-source image-recognition framework that it provides, with an interface that welcomes community contributions, means that it is highly adaptable. Indeed, with fairly modest additional effort, *ImageDataExtractor* could be extended to interrogate all sorts of ‘particles’ that are portrayed in 2-D images. Such extensions could involve highly disparate fields of scientific application, given that the tool could essentially be made relevant to any form of science that use 2-D images in their investigate armoury; from counting blood cells in medicine, to analysing biological specimens, to unlocking biochemical processes within organisms.

4. Glancing-angle PDF Analysis of DSSC Working Electrodes

This chapter outlines the development of a novel experimental technique called glancing-angle PDF (gaPDF), which allows PDF analysis of samples in the form of thin films. It also describes the experimental application of this technique to working electrodes of dye-sensitised solar cells (DSSCs) and the afforded conclusions regarding this interfacial structure. This chapter is based on a paper being prepared for publication.

My PhD supervisor, Jacqueline Cole, and I conceived and designed the research. Philip Chater of the Diamond Light Source, Jacqueline Cole and I developed the experimental methodology and led the experiments, with assistance from Othman Al Bahri and Apoorv Jain on the final experiment. Philip Chater performed the data reduction. Philip Chater and I conducted the data analysis with assistance from Jacqueline Cole. All density functional theory calculations were performed by Leon Devereux. Jacqueline Cole and I interpreted the results.

4.1 Summary

The efficient transport of electrons from the sunlight-harvesting dye molecules into the electrical circuit of a dye-sensitised solar cell (DSSC) is imperative to its effective operation. A dye···semiconductor interface comprises the working electrode of a DSSC. Dye molecules adsorb onto the semiconductor surface, whereupon they transfer electronic charge into the conduction band of the semiconductor; this process initiates the electrical circuit. It is therefore important to characterise this interfacial structure in order to understand how efficiently the dye binds, or anchors, onto the semiconductor surface and imparts charge transfer to it. Armed with such knowledge, the performance of DSSCs may then be improved systematically. The structural determination of a thin-film interface is nonetheless a challenging task. We herein report the results of a glancing-angle pair distribution function (gaPDF) experiment that generated synchrotron X-ray diffraction patterns of **N3**- and **N749**-sensitised DSSC working electrodes.

This gaPDF experimental approach represents the first diffraction-based strategy for the characterisation of intact DSSC working electrodes. The gaPDF structural signatures were compared with PDFs simulated from interfacial structures that were computed using density functional theory (DFT); these simulated structures showed the dyes anchoring in two distinct modes: a bridging bidentate and a monodentate ester configuration. The differences between the experimental observation and these simulated structures revealed a preference for each dye, **N3** and **N749**, to adopt a bidentate bridging dye anchoring mode when sensitised onto TiO_2. This work not only demonstrates the successful application of the gaPDF method to DSSC research, it also advocates the applicability of gaPDF to many types of thin-film samples.

4.2 Introduction

Dye-sensitised solar cells (DSSCs) represent promise for next-generation photovoltaic technology.[20] In particular, their transparency stimulates their development in solar-powered windows [21,22] for energy-sustainable buildings in future cities. In addition to representing the only truly transparent building-integrated photovoltaic technology, DSSCs excel in poor or ambient sunlight conditions. This stands in sharp contrast to other photovoltaic devices.[23,24] DSSCs also hold prospects as passive energy harvesters in wearable devices and textiles, [128-131] given their ability to be manufactured on flexible substrates or as fibres. More generally, the fact that DSSCs can be fabricated from cheap raw materials and undergo straightforward manufacturing processes means that their price-to-performance ratio exceeds that of fossil fuels, thus rendering DSSCs commercially competitive with more traditional energy sources. [132]

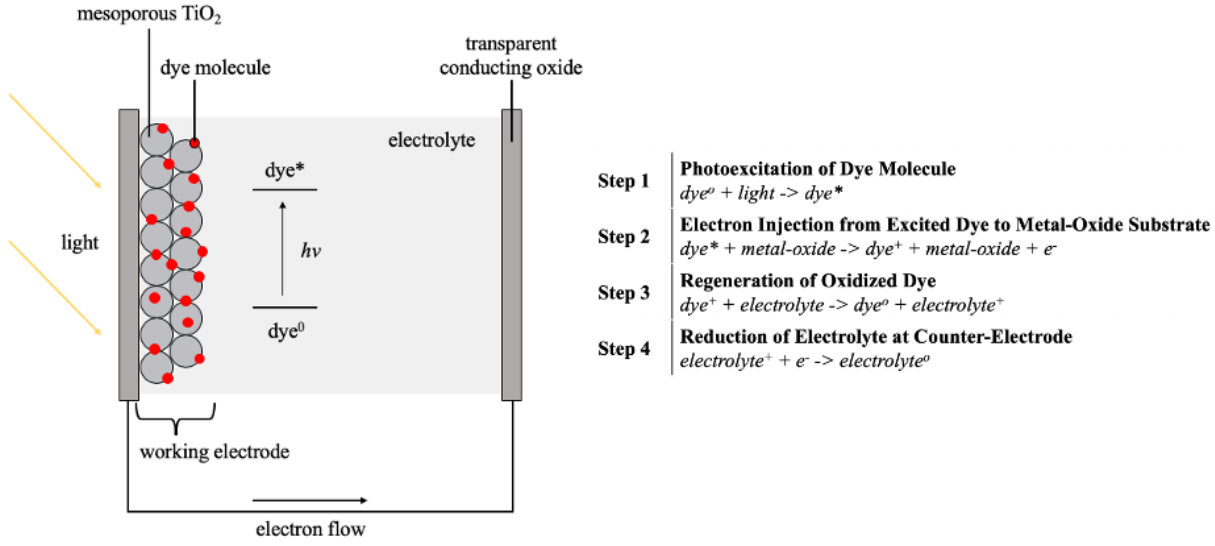


Figure 4.1 *General operational mechanism of DSSCs (left) and the four elementary steps that complete the circuit in DSSCs (right).*

The working electrode of a DSSC device usually adopts the form of a thin film whose thickness is typically of the order of 5-10 μm, depending on the manufacturing method. This film comprises a dye···semiconductor composite. The semiconductor is typically an oxide, most commonly, TiO_2 (anatase); as such, TiO_2 will be assumed to be the semiconductor for the rest of this chapter. The composite is prepared by depositing TiO_2 onto transparent conducting glass plates, and then sensitizing the resulting TiO_2 substrates for 24 hours with sunlight-harvesting dye molecules (Figure 4.1).

The adsorption of dye molecules onto the TiO_2 substrate is generally achieved via certain chemical substituents of the dye, which can act as anchoring groups to bind the dye molecules to the TiO_2 surface. These substituents are often carboxylic or cyanoacrylic acid groups.[133] The anchoring groups also facilitate the transfer of electrons from the photo-excited dye into the conduction band (CB) of the semiconductor, which initiates the electrical circuit. The performance of the anchoring group is therefore fundamental to the function of the DSSC, both in terms of its capacity of physical adsorption and in terms of its electron transfer characteristics. The structural determination of a dye···TiO_2 interface, including a resolution of its preferred

anchoring mode configuration, is thus essential for the strategic improvement of DSSC technology. [25-27]

The structural nature of dye···TiO_2 interfaces has already been investigated using various materials characterisation methods. For example, vibrational spectroscopy (infra-red and Raman) has been used to identify changes in dye···TiO_2 binding strengths or anchoring configurations via analysis of associated bond stretching or bending mode modulations.[28-31] X-ray photoelectron spectroscopy can detect if a dye is anchoring to a TiO_2 surface; if so, it can also quantify the corresponding dye···TiO_2 binding strengths.[7,32,33] X-ray absorption spectroscopy can identify oxidation states in dyes adsorbed onto TiO_2 surfaces. [33] Optical metrology, such as UV-vis absorption and emission spectroscopy, can be used to survey the propensity for a dye to aggregate on TiO_2 surfaces and determine if the dyes form H- or J-aggregates. [31,34] X-ray reflectometry can enumerate the thickness of the dye on a TiO_2 surface.[8,35-38] Neutron reflectometry can determine dye-layer thicknesses of a dye···TiO_2 interfacial structure while this DSSC working electrode is assembled with the device environment. [39] Concerted scanning tunnelling microscopy and scanning tunnelling spectroscopy (STM/STS) techniques have been used to demonstrate that multiple dye anchoring configurations can co-exist on a TiO_2 surface.[40-41] Atomic force microscopy (AFM) can be used to elucidate the spatial distribution of dye molecules on a TiO_2 surface. Accordingly, the extent to which dyes may aggregate on a TiO_2 surface can be estimated from AFM. [8,42,43] Complementary calculations using DFT and time-dependent (TD)-DFT calculations [27,134] are commonly performed. These calculations may be carried out using a dye···TiO_2 interface, or just the dye molecule, as the structural model. Such calculations are not only important for corroborating experimental findings; they can also provide uniquely accessible data about a material to help characterise its molecular properties, especially regarding its quantum energy levels and thermodynamics. Single-crystal X-ray diffraction has also been used to realise high-definition structural features about the dye molecules themselves. [135-137]

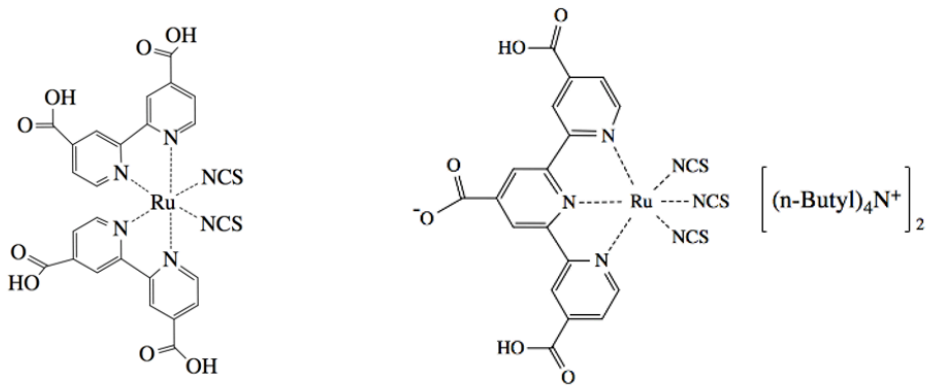
Despite this plethora of materials characterisation efforts, the DSSC field still lacks a consistent, quantitative and reliable method to probe the local atomic structure of dyes on TiO_2. Local atomic structure provides a geometric measure of the short-range order held within a material. A structural quantification of the short-range order contained within dye···TiO_2 interfaces would be particularly valuable with regards to unravelling the binding configuration of dye molecules that adsorb onto TiO_2 surfaces. This is especially pertinent in light of the recent STM/STS studies [40,41] which show that dye molecules can exist in multiple dye conformations on certain dye···TiO_2 interfacial structures; yet their atomic resolution could not be established by STM imaging. Several recent theoretical studies[31,138] have also corroborated this notion that more than one type of dye···TiO_2 binding configuration may co-exist. This evidence of dye-conformation multiplicity is an important finding since it presents fundamental implications for much of the modelling and experimental materials characterisation work done to date on dye···TiO_2 interfaces, whose analysis implicitly assumes one binding mode.

Moreover, structural data from a method that probes the short-range order of dyes on TiO_2 with atomic resolution would be highly complementary to the nanoscopically-resolved local structure of dye self-assemblies that has recently become available from grazing-incidence small-angle X-ray scattering (GI-SAXS) and AFM studies on DSSC working electrodes. [139,140] Indeed, dye-self assemblies on TiO_2 surfaces appear to exhibit short, medium and long-range order.

This study explores the feasibility of probing the local atomic structure of DSSC working electrodes with atomic pair distribution function (PDF) analysis. PDF analysis has become a powerful technique for the characterisation of nanostructures, [1] having been borne out of a field of materials characterisation on liquids [44-47] and glasses [44,48,49] that exclusively exhibit local atomic structure. PDF analysis has previously been applied to bulk powder constituents of perovskite solar cell working electrodes. There, the light-sensitive perovskite, methylammonium lead iodide ($MAPbI_3$), is held within a mesoporous TiO_2 matrix, whereby this composite is called "meso-$MAPbI_3$". Bulk powder samples were prepared by scraping off a sufficient amount of

meso-$MAPbI_3$ material from many thin film electrodes of perovskite solar cells. [50] The study revealed a two-component nanostructure and underscored the need for a full characterisation in order to systematically improve cell performance. PDF analysis has also successfully been used to investigate the structure of $FeSb_3$ thin films, using a technique called thin film PDF (tfPDF). [51] However, PDF analysis has not yet been used for the investigation of the thin-film working electrodes of DSSCs.

This study herewith investigates the structure of DSSC working electrodes, which comprises a dye···TiO_2 interface in the form of a thin film. We have a particular focus on determining local atomic structure associated with the dye anchoring modes on TiO_2 surfaces. We employ two dyes as case studies for this study, which are the quintessential ruthenium-based DSSC dyes, **N3** and **N749** (Black Dye). These dyes serve the DSSC field as an industry standard and to provide wide spectral capture, respectively. The molecular structures of *cis*-bis(isothiocyanato)bis(2,2′-bipyridyl-4,4′-dicarboxylato)ruthenium(II) (**N3**) and tris(*N*,*N*,*N*-tributyl-1-butanaminium) [[2,2″6′,2″-terpyridine]-4,4′,4″-tricarboxylato(3-)*N*1,*N*1′,*N*1″]tris(thiocyanato-N) hydrogen ruthenate (**N749**) are illustrated in Figure 4.2.



N3 **N749**

Figure 4 2 *Molecular structures of* ***N3*** *(left) and* ***N749*** *(right).*

N3 bears four carboxylic acid groups while **N749** furnishes three such substituents. Carboxylic acid groups are generally considered to provide the best anchoring groups for adsorbing dyes onto TiO_2 surfaces, [24,141] once they have been deprotonated to afford their carboxylate forms. However, the binding mode configuration(s) of these carboxylate ions on TiO_2 are contended for ruthenium-bipyridyl based dyes. Some optical [142] and vibrational[142-143] spectroscopy studies suggest that the dyes bind via the bidentate chelating or bidentate bridging mode shown in Figure 4.3, middle and right, respectively. Other optical and vibrational studies [142] argue that a monodentate ester binding mode (Figure 4.3, left) as well as bidentate bridging and chelating modes exist in dye···TiO_2 interfacial structures. Meanwhile, DFT calculations [144] report a strongly unfavourable adsorption energy associated with the bidentate chelating mode, suggesting that monodentate ester and bidentate bridging are the two possible anchoring modes. The aim of our study is to demonstrate that PDF analysis of DSSC working electrodes can be applied to dispel this contention by realizing direct evidence from the local atomic structure of dye···TiO_2 interfaces.

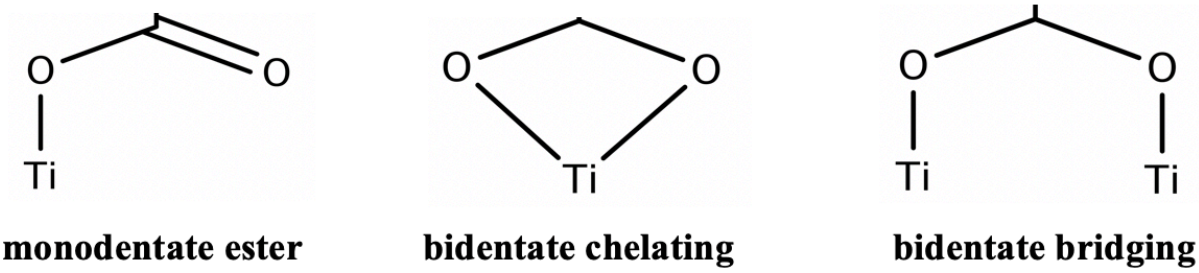


Figure 4.3 *Possible binding modes for carboxylate group anchors onto TiO_2.*

DSSC working electrodes containing **N3** or **N749** dye adsorbed onto mesoporous TiO_2 were examined via PDF analysis of synchrotron X-ray diffraction data collected at the beamline, I15-1, of the Diamond Light Source, UK, using a novel setup that we have termed glancing-

angle PDF (*gaPDF)* (Figure 4.4). Complementary density functional theory (DFT) calculations were employed to provide structural models of **N3**···TiO_2 and **N749**···TiO_2 interfaces where the dyes adsorb in both the monodentate ester and bidentate bridging mode. These structural models were also used to simulate PDFs of the corresponding dye···TiO_2 interfacial structures for the purposes of comparison.

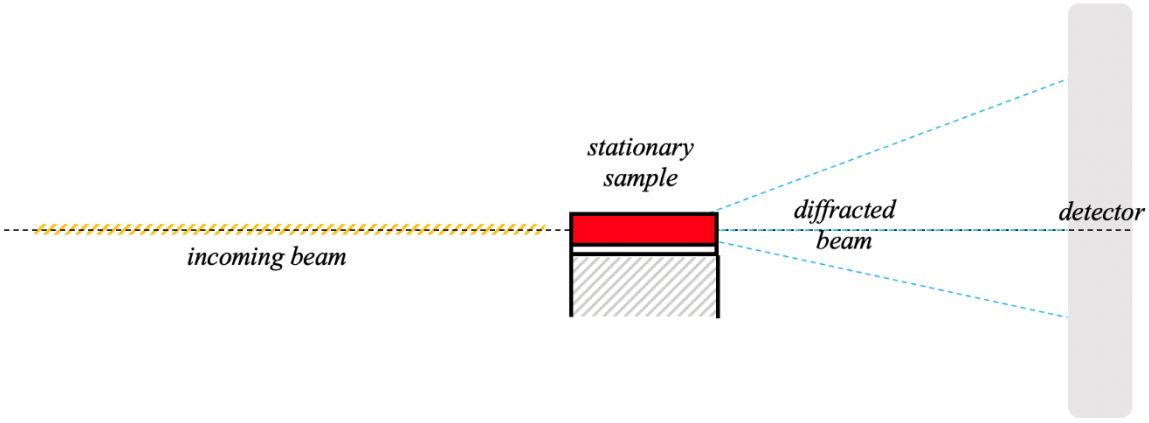


Figure 4.4 *Schematic illustration of the experimental setup for the glancing-angle PDF (gaPDF) analysis. The working electrode (dye sensitised TiO_2 thin film, shown in red) is deposited on a transparent conducting glass plate (white) then placed on a sample stage (shown in textured grey). The schematic is not to scale.*

4.3 Experimental Methods

Sample Preparation. The samples were prepared following traditional DSSC fabrication methods, up until the point before the addition of electrolyte and sealing with the counter electrode. Thereby, Delta Technologies CB-90IN-10X10/0.5 glass slides (10 x 10 x 0.5 mm^3 boro-aluminosilicate glass slide coated with indium tin oxide on one side) were cleaned in a James Products Sonic 4D 120W ultrasonic bath using an aqueous solution of Decon (5%; 15

min), deionised water (15 min), and ethanol (15 min). Then, Dyesol 18NR-T transparent titania paste (anatase nanoparticles) was doctor-bladed onto the conductive faces of the glass substrates, which were subsequently sintered (500 ºC; 30 min). Using the equipment in our laboratory, each application is expected to result in a layer of ~10 μm thickness. The doctor-blading step was repeated once more to increase the thickness of the TiO_2 layer to approximately ~20 μm. Unsensitised TiO_2 samples were ready at this stage. The remaining samples were divided and immersed (24 h; covered petri dishes) in 0.5 mM solutions (*t*-BuOH/MeCN, 1:1, v/v) of Solaronix Ruthenizer 535 (**N3**) or Solaronix Ruthenizer 620-1H3TBA (**N749**). The samples were then rinsed using pure acetonitrile and briefly dried under a flow of nitrogen.

gaPDF X-ray Diffraction Setup. Synchrotron-based X-ray diffraction data were collected at the Diamond Light Source XPDF beamline, I15-1, using a photon wavelength of $\lambda = 0.1611$ Å and a Perkin Elmer XRD 1611 CP3 area detector. An experimental setup for gaPDF is possible due to the vertically narrow focused beam of I15-1 (full width at half maximum at its focus: 9 μm). This setup allows placing only the thin film, which comprises the dye···TiO_2 interface, into the beam path, so that the working electrode can be observed in its intact form, rather than as a powder sample.

Such a refined experimental setup requires multiple calibration steps. Initially, a vertical calibration was carried out in order to ensure maximum sample exposure, while concomitantly excluding the substrate from the beam path. This was accomplished by recording multiple short exposures across a range of sample stage y-positions (starting from ±50 μm from the assumed centre position, narrowing the range down until the precision limit of ±5 μm was reached), and selecting the position that led to the highest overall scattering intensity under concomitant preservation of the resolution of the Bragg peaks. The beam was half-cut with the sample in order to exclude the substrate from the beam path. Subsequently, a rotational calibration was carried out. The sample was oriented in parallel to the incoming beam to ensure that the entire length of the sample was placed in the beam path; this results in increased scattering and

consistency across the samples. This was accomplished by carrying out another set of short exposure scans, this time across a small range of rotational positions relative to the z-axis, starting at the assumed parallel position. These two steps were repeated with increasingly precise scans. Figure 4.5 illustrates the aforementioned steps of the calibration process.

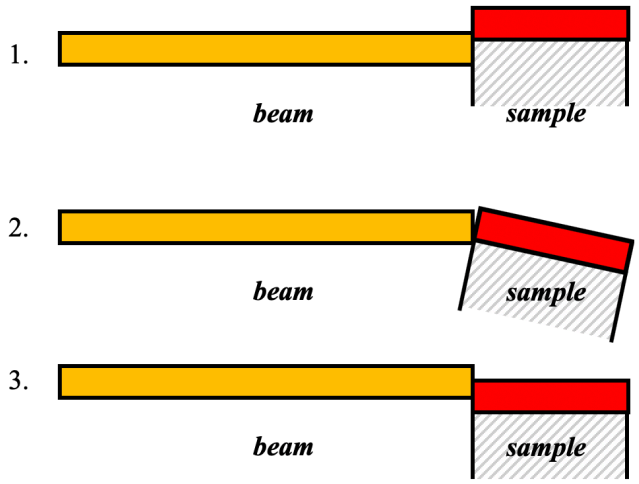


Figure 4.5 *(1) Improper vertical alignment; (2) improper rotational alignment; (3) optimum alignment.*

The subtraction of background contributions in the gaPDF setup is non-trivial. As the alignment process was carried out in reference to the thin film (working electrode), the exact relative position of the substrate (glass slide) is unknown. Therefore, background substrate measurements were taken over a range of vertical positions, starting from half-cutting the beam (maximum expected substrate scattering) to further away than the maximum expected thickness for the film (~20 μm; determined from a visual analysis of the scans). The most accurate background measurement was determined by visually comparing the raw diffraction intensity I_{exp} of each measurement to that of the sample and selecting the best fitting one.

I_{exp} data were processed using the Gudrun package [145] while experimental and simulated PDFs, *D(r)*, were calculated using Gudrun [145] and Topas [146]. We use *D(r),* as defined by Keen, [93] which in effect describes the average number density of atoms at a given interatomic distance *r,* according to:

$$i(Q) = \frac{I_{exp} - \langle f^2 \rangle}{\langle f \rangle^2} \tag{4.1}$$

$$D(r) = \frac{2}{\pi} \int_0^\infty Q i(Q) \sin(Qr)\, dQ \tag{4.2}$$

Fourier transforms were performed at various values of Q_{max} in order to understand the extent of artefacts that might manifest in *D(r)* owing to Fourier series termination effects. The *D(r)* presented below is from the most suitable value of Q_{max} = 25 Å^{-1}. It can be seen in the full range of Q_{max} values tested (provided in Appendix A1) that artefacts owing to Fourier series termination effects dissipated >1.4 Å.

Conventional PDF X-ray Diffraction Setup for Control Experiments. This study seeks to identify the anchoring modes in these dye···TiO_2 interfacial structures. Therefore, contributions to the gaPDF scattering pattern that arise specifically from dye molecules are needed as a reference.

PDF data on these pure dyes were collected in their bulk material form as part of a mail-in experiment during the commissioning of beamline I15-1 using a spinning capillary setup. Solaronix 535 (**N3**) dye was loaded into a 1 mm quartz glass capillary and measurements were performed using a photon wavelength of λ = 0.1583 Å. Using the Gudrun package [145], the 2-D data were corrected for polarisation and flat-field, then integrated to 1D and processed into PDFs using Q_{min} = 1.0 Å^{-1} and Q_{max} = 23.5 Å^{-1}. A Lorch Function [147] (Δ_1 = 0.1) was applied to suppress spurious low-r features.

Density Functional Theory Calculations. The scattering contributions of dye molecules to the gaPDF scattering pattern are given by the aforementioned conventional PDF experiments. However, these contributions from measurements on the pure dyes naturally exclude any dye···TiO_2 interactions. Dye···TiO_2 interfacial structures need to be simulated in order to obtain the closest comparison of the gaPDF data. Accordingly, geometrically-optimised structures of **N3** and **N749** on the [101] surface of anatase TiO_2 were computed using DFT, within the NWChem computational chemistry package.[148] All calculations employed the B3LYP functional [149,150] and the 3-21G(d) basis set for elements H, C, N, O and S, while the LANL2DZ effective core potential and basis set were used for Ti and Ru.[151]

A geometric model of the **N3** molecule and a $(TiO_2)_{38}$ slab *in vacuo* was initially optimised. This size of TiO_2 slab was selected since other studies [152-154] have shown that it is large enough to allow a scientifically realistic simulation while maintaining computational efficiency. Then, one anchoring COOH group of **N3** was deprotonated and anchored onto the [101] surface of the TiO_2 slab via a bidentate-bridging (BB) mode (Figure 4.2, right) to five-coordinate Ti atoms. The surplus H^+ that results from this adsorption process was bound to a two-coordinate oxygen atom on the TiO_2 surface for charge-balancing. The compound-slab-adsorbate structure was optimised until all forces on atoms were below 0.023 eV $Å^{-1}$ and the root-mean-square (RMS) force was below 0.015 eV $Å^{-1}$. The **N3**···TiO_2 interfacial structure afforded from this optimisation process (Figure 4.6) featured two anchoring modes: bidentate-bridging (BB) via the carboxylate ion, which has been generated via the simulated deprotonation of **N3** in this calculation, and monodentate binding via a carboxyl group in the carboxylic acid of **N3** that shares the same bipyridyl ligand as the carboxylate ion.

The starting structure for the adsorbed **N749** model was built by modifying the final structure of the adsorbed **N3** dye. This was subsequently optimised until the same convergence criteria were met.

The geometries of these DFT-generated dye···TiO_2 structural models were imported into Topas [146] in order to simulate *D(r)* and thence *ΔD(r)* profiles that describe a theoretical model of a PDF that pertains to a structural configuration utilizing both possible binding modes (BB and monodentate ester). Appropriate peak widths for the PDF simulations were determined from observed PDF data of the pure dyes. This allowed accurate comparisons with the data, which had to be ensured given the extended beam path (1 cm) imposed by the sample geometry. These models are shown in Figure 4.7, together with the experimentally-generated PDF profiles, for the purposes of comparison.

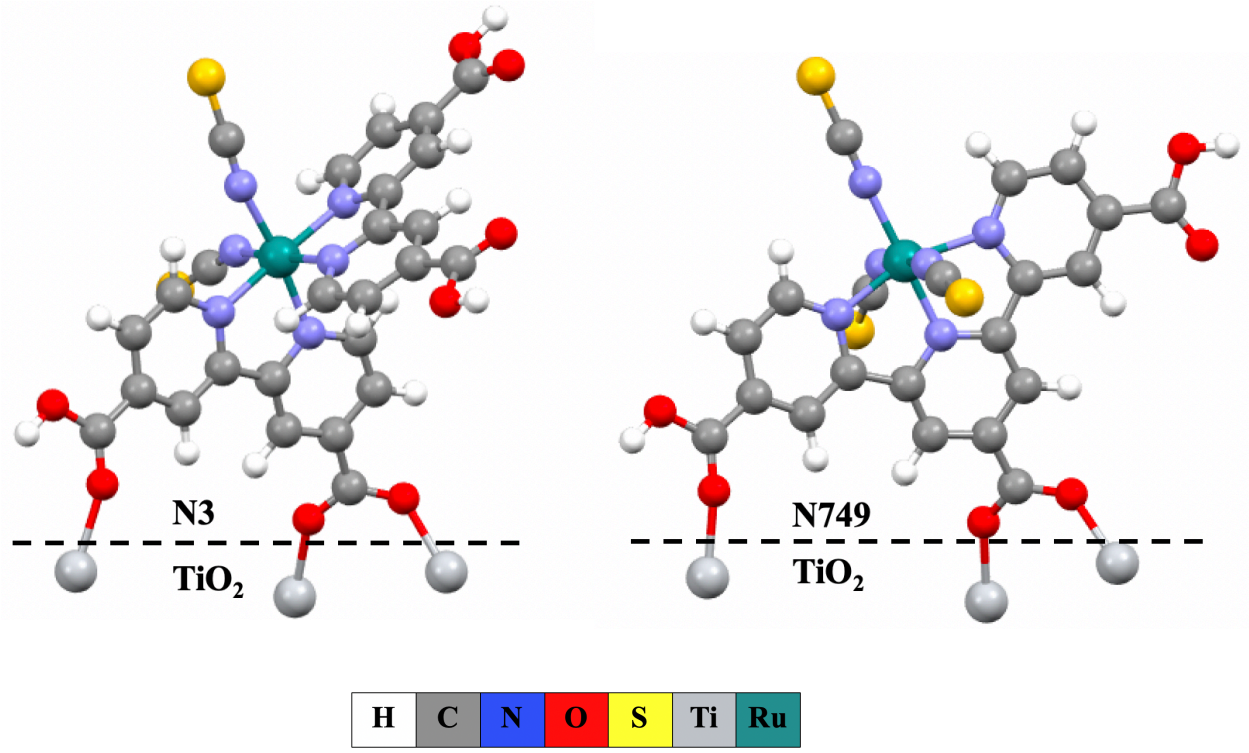


Figure 4.6 *Optimised structures of **N3**···$(TiO_2)_{38}$ (left) and **N749**···$(TiO_2)_{38}$ (right). Only the titanium atoms that the dyes are anchored on were included in the calculation of the simulated PDF. Therefore, only these are shown in the diagrams above.*

4.4 Results and Discussion

The ΔPDFs [$\Delta D(r) = D(r)^{exp}_{sensitized} - D(r)^{exp}_{unsensitized}$] yield the background-adjusted subtraction of data on unsensitised TiO_2 films from those of the dye-sensitised TiO_2 films. These represent the structural features of the dye···TiO_2 interface and the dye molecules. Figure 4.7 shows these ΔPDFs, together with the simulated $D(r)^{sim}$ profiles of the corresponding dye molecules on TiO_2.

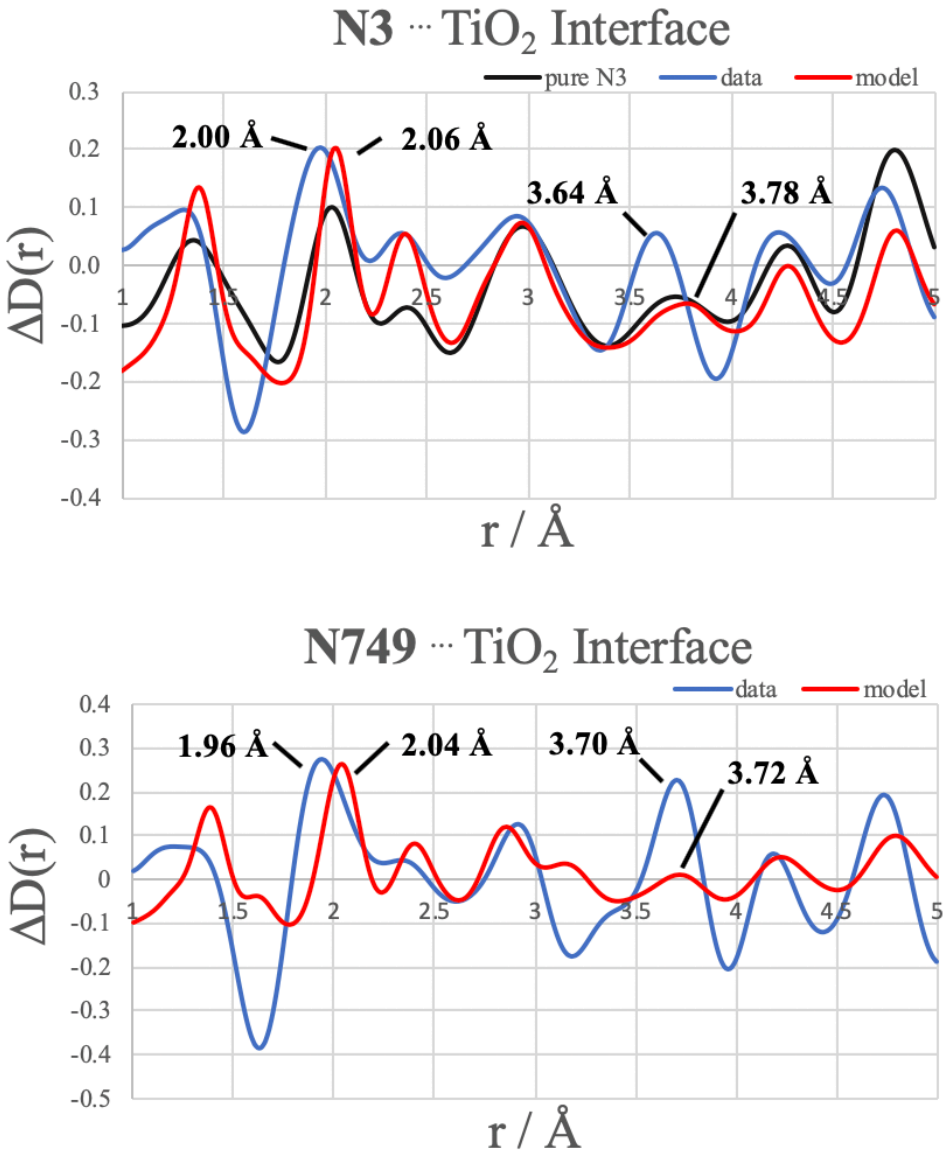


Figure 4.7 *(Red) Simulated and (blue) experimental ΔD(r) profiles representing the dye···TiO_2 interface of* ***N3****- and* ***N749****-sensitised TiO_2 substrates. (Black) Normalised D(r) of pure* ***N3*** *dye.*

Figure 4.7 (top) also shows the normalised *D(r)* signature of the pure **N3** dye, collected in a capillary setup during a previous experimental run at beamline I15-1. There is good agreement between features in r, for the results on the dye···TiO_2 interfaces from the gaPDF experimental setup and this independent measurement on the pure dye using a standard capillary setup. This

level of agreement evidences good overall validity of the simulated PDF as well as the gaPDF experimental setup for the dye···TiO_2 interfacial measurements. The 1-5 Å range displayed is where the anchoring interactions are expected to reside.

Statistical Analysis of the Obtained Results. Pearson correlation coefficients were calculated for **N3**- (0.55) and **N749**-sensitised TiO_2 (0.47) between 1 and 5 Å in order to quantify the agreement between the experimental and simulated $\Delta D(r)$. Such coefficients can range from -1 to 1 (-1: perfect anti-correlation; 0: no correlation; 1: perfect correlation).[95] Figures 4.2 and 4.3 show that the model structures of both dyes are quite similar, as one might expect given their similar dye···TiO_2 interfacial structures calculated from DFT, that are shown in Figure 4.6. This good level of similarity is corroborated by the comparable Pearson correlation coefficients (0.93) calculated over the same range of r, using the simulated $D(r)^{sim}$ from each model. The comparable value calculated using $\Delta D(r)^{exp}$ also indicates good internal agreement in the experimental data (0.87), which is indicative of good reproducibility. This level of similarity should be even more apparent over the range of r that relates to the direct dye···TiO_2 interactions, since both dyes are expected to anchor using either the bidentate bridging or the monodentate ester mode (Figure 4.2), or both of them. This is indeed the case: between 1.9 and 3 Å, the simulated $D(r)^{sim}$ from the two dye models compare favourably with a Pearson correlation coefficient of 0.91. The corresponding value from the comparison of the cognate experimental ΔPDFs shows an even better agreement, with a Pearson correlation coefficient of 0.94. These high Pearson correlation coefficients are strong indicators of good quality experimental data; they also offer promise to the success of the gaPDF set up. However, the Pearson coefficients for the experiment-minus-model data pairs are markedly lower over the same range (1.9-3 Å; **N3**: 0.65; **N749**: 0.73). Given that a high robustness of the experimental data has been assured, this poorer agreement indicates that there may be imperfections in the model.

Interpretation of the Results. It should be noted here that a perfect correlation between the experimental data and the theoretical model is not expected, as it is precisely the deviations from the simulated structure that we set out to investigate. The moderate level of correlation of $\Delta D(r)$

between experiment and simulation over the full range 1-5 Å, given by the Pearson correlation coefficients, 0.55 (**N3**-sensitised TiO_2) and 0.47 (**N749**-sensitised TiO_2), suggests that: (a) the gaPDF technique and associated experimental setup is indeed able to observe the dye···TiO_2 interface; and (b) the simulated models of these interfacial structures requires a modest level of refinement, perhaps to consider the fuller range of options for anchoring modes. However, certain inferences can still be made from these differences using the current models.

This process of inference is aided by first considering which pairwise atomic correlations can contribute to the experimental and simulated $\Delta D(r)$ profiles. Those contributing to the observed experimental $\Delta D(r)^{exp}$ arise from the direct dye···TiO_2 interactions responsible for absorption, bonds within the dye molecules themselves, dye···dye interactions and n^{th} neighbour dye···TiO_2 interactions. The $D(r)^{sim}$ profiles from the simulated models are also able to emulate direct dye···TiO_2 interactions and bonds within the dye molecules themselves, but they cannot account dye···dye interactions and n^{th} neighbour dye···TiO_2 interactions. These absences in the model arise because the model was simulated using only a single dye molecule on a TiO_2 slab, whereby most of the $(TiO_2)_{38}$ cluster was removed to prevent the $D(r)^{sim}$ signal from being overwhelmed by the larger TiO_2 slab.

However, dye···dye interactions and n^{th} neighbour dye···TiO_2 interactions are not expected to manifest < 3 Å, where the contributions from the adsorbing bonds can be observed. Moreover, the core structure of each dye molecule is not expected to change, so bonds within the dye molecules themselves are most likely irrelevant to the $\Delta D(r)$. Thus, the deviations between the experimental and simulated $\Delta D(r)$ below 3 Å should be attributed exclusively to anchoring behaviour that has not been modelled. As mentioned before, the simulated model features one instance each of the two anchoring modes (BB and monodentate ester) per dye molecule (Figure 4.8).

Comparing Competing Binding Modes. The strongest contributions from the anchoring groups are expected to be the average of the relevant direct dye···TiO_2 interactions (O_{dye}···Ti) that are

shown in Figure 4.8 (**N3**: 2.07 Å; **N749**: 2.02 Å). These lengths correspond to the second peak in each $\Delta D(r)$. Pairwise atomic correlations from these direct dye···TiO_2 interactions are thus likely to be the cause deviations between experimental data and the simulation within this peak at ~2 Å. Indeed, this second peak in each $\Delta D(r)$ is the location in r that exhibits the most striking difference between the experimental data and simulated model, *i.e.*, a shift to shorter distances from the simulation to the experiment for both samples (**N3**: 2.06 Å → 2.00 Å; **N749**: 2.04 Å → 1.96 Å). This is attributed to the shorter-than-expected average length of the direct dye···TiO_2 interaction in the experimental $\Delta D(r)^{exp}$, which suggests a prevalence of the BB mode, *i.e.*, the mode with shorter direct dye···TiO_2 interactions.

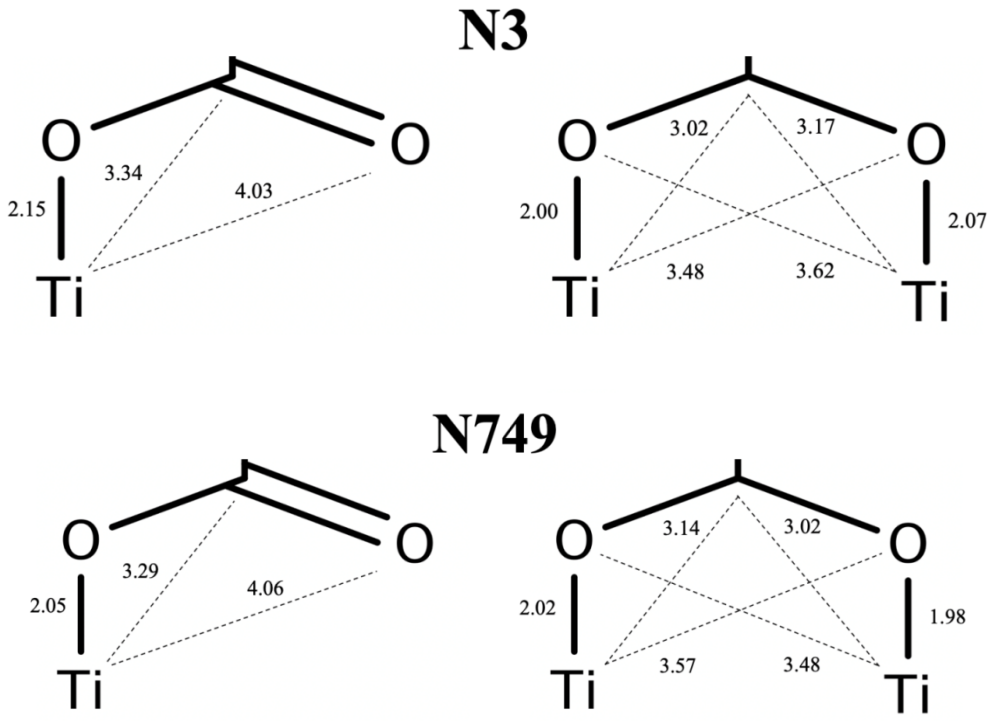


Figure 4.8 *Relevant carboxylic group bond lengths (in Å) associated with monodentate ester (left) and bidentate bridging (right) anchoring modes as calculated in the models of **N3** and **N749** on* $(TiO_2)_{38}$.

The other most striking differences between the experimental data and simulated model occur at the peaks centred at ~3.6 Å, a separation that is also associated with the second neighbour $O_{dye}\cdots Ti$ interaction in the bidentate bridging mode (see Figure 4.8). At this separation, signal appears much stronger in the experimental data than in the model, for both dyes, implying a higher abundance of pairwise atomic correlations than expected in the model. The second neighbour $O_{dye}\cdots Ti$ interaction expected in a monodentate ester binding mode, lies at ~4 Å and no discrepancy between the data and model at this separation is observed in the *D(r)*. These results thus suggest a preference for the BB mode, which is consistent with attribution of the peak at 2 Å to BB anchoring correlations as per the findings discussed above . These deductions are aligned with other research on these dyes, where optical spectroscopy, vibrational spectroscopy and DFT likewise suggest a relative preference for the bidentate bridging mode. [26,142-144]

4.5 Conclusions

The thin-film working electrodes of **N3-** and **N749-**sensitised DSSCs, comprising a dye$\cdots TiO_2$ interface, were probed using PDF analysis of synchrotron-based X-ray diffraction data. The thin-film nature of this study required the development of a novel experimental setup for PDF data acquisition, glancing-angle PDF (gaPDF). Analogous data were collected on unsensitised DSSCs, allowing us to employ ΔPDF analysis to isolate the interfacial structure, by subtracting out PDF contributions from the bare TiO_2 substrate. These $\Delta D(r)^{exp}$ profiles were compared to $\Delta D(r)^{sim}$, which were generated from simulated models of the interfacial structure, using DFT calculations. Pearson correlation coefficients between experimental $\Delta D(r)$ and simulated $\Delta D(r)$ were only moderate, implying that although it is possible to observe the interface from the data, our models require refinement. Closer examination of the deviations of the experimental data from the models implied shorter-than-expected average bond lengths at values of r associated with anchoring groups. This suggests a prevalence of the BB mode since it features shorter dye$\cdots TiO_2$ interactions relative to the alternative monodentate ester binding mode. This study demonstrates the successful application of the gaPDF method to DSSCs and underscores the need of further investigation into DSSC anchoring groups.

5. Gd···Gd Separations in Amorphous $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$

PDF and statistical analysis were performed on data gathered from an anomalous X-ray scattering (AXS) study on a Gd-doped rare-earth phosphate glass (REPG) with composition $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$. The aim was to investigate its local structure with a specific interest in the closest Gd···Gd separations. This chapter is based on a paper published in *Physical Review Materials* [100] (see *Publications* on page II). All figures and tables were reproduced in accordance with the guidelines set by the publisher.

My contribution to this project was the analysis of ΔPDFs and the subsequent application of various statistical techniques, which allowed various structural conclusions to be drawn, including the sought-after Gd···Gd separations.

5.1 Summary

A Gd K-edge anomalous X-ray scattering (AXS) study is performed on the rare-earth (R) phosphate glass, $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$, in order to determine Gd···Gd separations in its local structure. The minimum rare-earth separation is of particular interest given that the optical properties of these glasses can quench when rare-earth ions become too close to each other. To this end, a weak Gd···Gd pairwise correlation is located at 4.2(1) Å which is representative of a meta-phosphate R···R separation. More intense first neighbour Gd···Gd pairwise correlations are found at the larger radial distributions, 4.8(1) Å, 5.1(1) Å and 5.4(1) Å. These reflect a mixed ultra-phosphate and meta-phosphate structural character, respectively. A second neighbour Gd···Gd pairwise correlation lies at 6.6(1) Å which is indicative of meta-phosphate structures. Meta- and ultra-phosphate classifications are made by comparing the R···R separations against those of rare-earth phosphate crystal structures, $R(PO_3)_3$ and RP_5O_{14} respectively, or difference pair distribution function (ΔPDF) features determined on similar glasses using difference neutron scattering methods. The local structure of this glass is therefore found to display multiple rare-

earth ion environments, presumably because its composition lies between these two stoichiometric formulae. These Gd···Gd separations are well resolved in the ΔPDFs that represent the AXS signal. Indeed, the spatial resolution is so good that it also enables the identification of R···X (X = R, P, O) pairwise correlations up to r ~ 9 Å; their average separations lie at r ~ 7.1(1) Å, 7.6(1) Å 7.9(1) Å, 8.4(1) Å and 8.7(1) Å. This is the first report of a Gd K-edge AXS study on an amorphous material. Its demonstrated ability to characterise the local structure of a glass up to such a long-range of r, heralds exciting prospects for AXS studies on other ternary non-crystalline materials. However, the technical challenge of such an experiment should not be underestimated, as is highlighted in this work where probing AXS signal near the Gd K-edge is found to produce inelastic X-ray scattering that precludes the normal AXS methods of data processing. Nonetheless, it is shown that AXS results are not only tractable, but they also reveal local structure of rare-earth phosphate glasses that is important from a materials-centred perspective and which could not be obtained by other materials characterisation methods.

5.2 Introduction

REPGs show great promise for applications in the laser and optoelectronics industry,[52-56] as rare-earth ions exhibit the required energy levels for successful population inversion, and as the nonlinear refractive index of the phosphate glass is sufficiently high to generate the desired effects without causing beam breakup. Particularly interesting are REPGs of the type $(R_2O_3)_x(P_2O_5)_{1-x}$ (x = 0.167 (ultraphosphate) to x = 0.25 (metaphosphate)). REPGs in this stoichiometric range exhibit various attractive properties at low temperatures (T < 20 K), including acoustic mode softening, which results in negative thermal expansion coefficients [57] and bulk moduli,[58] as well as unprecedented magnetic, optomagnetic, and optoacoustic phenomena.[59]

A proper understanding the structure-property relationships of REPGs would help to significantly advance research and development in this area. The atomic structures of $(R_2O_3)_x(P_2O_5)_{1-x}$ (x = 0.167−0.250; R = La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, and Er) have

been probed by a variety of spectroscopy and diffraction techniques, including conventional X-ray[60-67] and neutron[61,63,65,68] diffraction, as well as rare-earth K-edge[69] and L_{III}-edge[60,70-72] X-ray absorption spectroscopy (XAS), ^{27}Al and ^{31}P solid-state NMR,[62,68] and Fourier transform infrared spectroscopy (FTIR).[73] The combination of results from these previous studies has afforded a model of the local atomic structure of such REPGs, comprehensive only out to r ~ 3.8 Å. However, the closest R···R separations cannot be discerned by these conventional techniques, implying this separation must lie at r > 3.8 Å. This structural feature is particularly important to define, given that rare-earth ion clustering can inhibit certain optical properties through quenching.

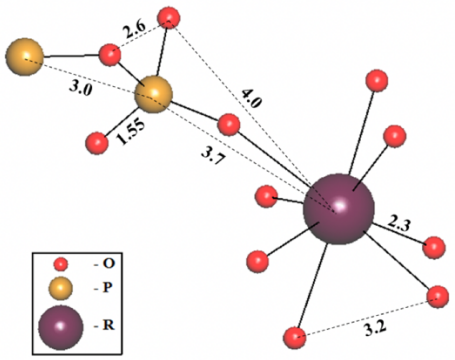


Figure 5.1 *Local structure of rare-earth (R) phosphate glasses, $(R_2O_3)_x(P_2O_5)_{1-x}$, showing its modelled geometry up to a radial distribution of r = 3.8 Å.*

5.3 AXS Method

The anomalous X-ray scattering technique relies on wavelength-dependent real (f'(λ)) and imaginary (f''(λ)) terms of the atomic scattering factor, f, presenting anomalous scattering contributions to the overall X-ray scattering when diffraction studies are conducted in the vicinity of the X-ray absorption edge of one of the elements (R) in a sample under study. These anomalous scattering contributions augment the wavelength-independent Thomson scattering factor, f_0, to afford the overall atomic scattering factor for X-ray diffraction according to

$$f_R = f_0 + f'(\lambda) + if''(\lambda) \quad (5.1)$$

The difference between X-ray diffraction data collected at two X-ray energies, one on and one far from the X-ray absorption edge of the element of interest (R) within a sample, affords exclusively the anomalous scattering contributions of R. For an amorphous material, these are present as pairwise R···X (X = any element) correlations.

Moreover, f' and f'' are related via a Kramers-Kronig transformation: [155,156]

$$f' = \frac{2}{\pi}\int_0^\infty \frac{E' f''(E')}{E^2 - E'^2} dE' \quad (5.2)$$

This shows that the largest f' values will be obtained in cases where the energy difference between the pair of X-ray diffraction measurements that comprise an AXS experiment is maximised. X-ray absorption K-edges are therefore favoured for AXS experiments as they offer the highest possible energy for an X-ray absorption edge of a given element, assuming that there is no interference from other X-ray absorption edges in this energy regime.

As contributions from the Gd atoms in the sample $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$ were of specific interest, the AXS method was deemed particularly suitable. The nature of this method affords ΔPDFs by subtracting a data set with an X-ray energy far from the Gd K-edge from that of one on the Gd K-edge. Thus, the ΔPDFs theoretically feature exclusively Gd···X (X = any other element in the sample) atomic pairwise correlations.

As with any ΔPDF study, the relatively low signal-to-noise ratio was a challenge that required resolving. In this aggravated case, contributions from all other (non-Gd···X) atomic pairs have to be disregarded along with the experimental noise.

My contribution to the project was to calculate and analyse statistical correlations between all possible pairs of ΔPDFs using three figures of merit, i.e., the Pearson, Spearman, and Kendall rank coefficients [95] in order to reliably isolate the Gd···Gd signal.

5.4 AXS Experiment

The $(Gd_2O_3)_{0.232}(P_2O_5)_{0.768}$ sample was synthesized using the method described by Mierzejewski *et al*. [157]

A disk (thickness: 0.5 mm) of finely powdered $(Gd_2O_3)_{0.232}(P_2O_5)_{0.768}$ was held within an aluminium annulus and between two Kapton windows; the assembly was mounted onto a six-circle diffractometer in flat-plate geometry. This diffractometer was housed on the synchrotron beam line 1-BM at the Advanced Photon Source, Argonne National Laboratory, IL, USA. X-rays were delivered to the instrument via a Si(400) crystal monochromator and detected using an energy-dispersive germanium detector. X-ray slits were optimised to afford a 0.1 $Å^{-1}$ resolution with a usable dynamic range (Q = 0.4 - 30 $Å^{-1}$). Diffraction scans were performed using 2θ/θ geometry, at X-ray energies of 10 eV, 40 eV, 300 eV, and 500 eV below the gadolinium K-edge (50.239 keV) which was calibrated using a gadolinium foil.

5.5 Data Analysis

ΔPDFs were calculated using all possible combinations of the four *T(r)* (K-10 eV, K-40 eV, K-300 eV, K-500 eV) functions provided by my collaborators. Such pairings enabled a validation of the self-consistency of the AXS signal within the PDF. For example, the largest AXS contribution should arise in a ΔPDF produced by the subtraction of a PDF associated with an X-ray energy far from the Gd K-edge from the one whose energy lies closest to this edge, *i.e.*, maximizing $\Delta f(\lambda)$. Given the four X-ray energies where data were acquired, PDF[(K-300 eV) – (K-10 eV)] and PDF[(K-500 eV) – (K-10 eV)] should show the largest AXS contributions and

thus the most noticeable R···X correlations; these two difference pairs should show comparable results given that 300 eV and 500 eV are sufficiently far from the Gd K-edge.

Conversely, the ΔPDF corresponding to PDF[(K-500 eV) – (K-300 eV)] should essentially feature 'noise', given that both data sets were collected at energies far from the Gd K-edge, *i.e.*, $\Delta f(\lambda) = 0$. ΔPDFs generated from PDF[(K-300 eV) – (K-40 eV)] and PDF[(K-500 eV) – (K-40 eV)] should show comparable results, with an AXS signal that is less intense than that of cognate differences involving the K−10eV data set.

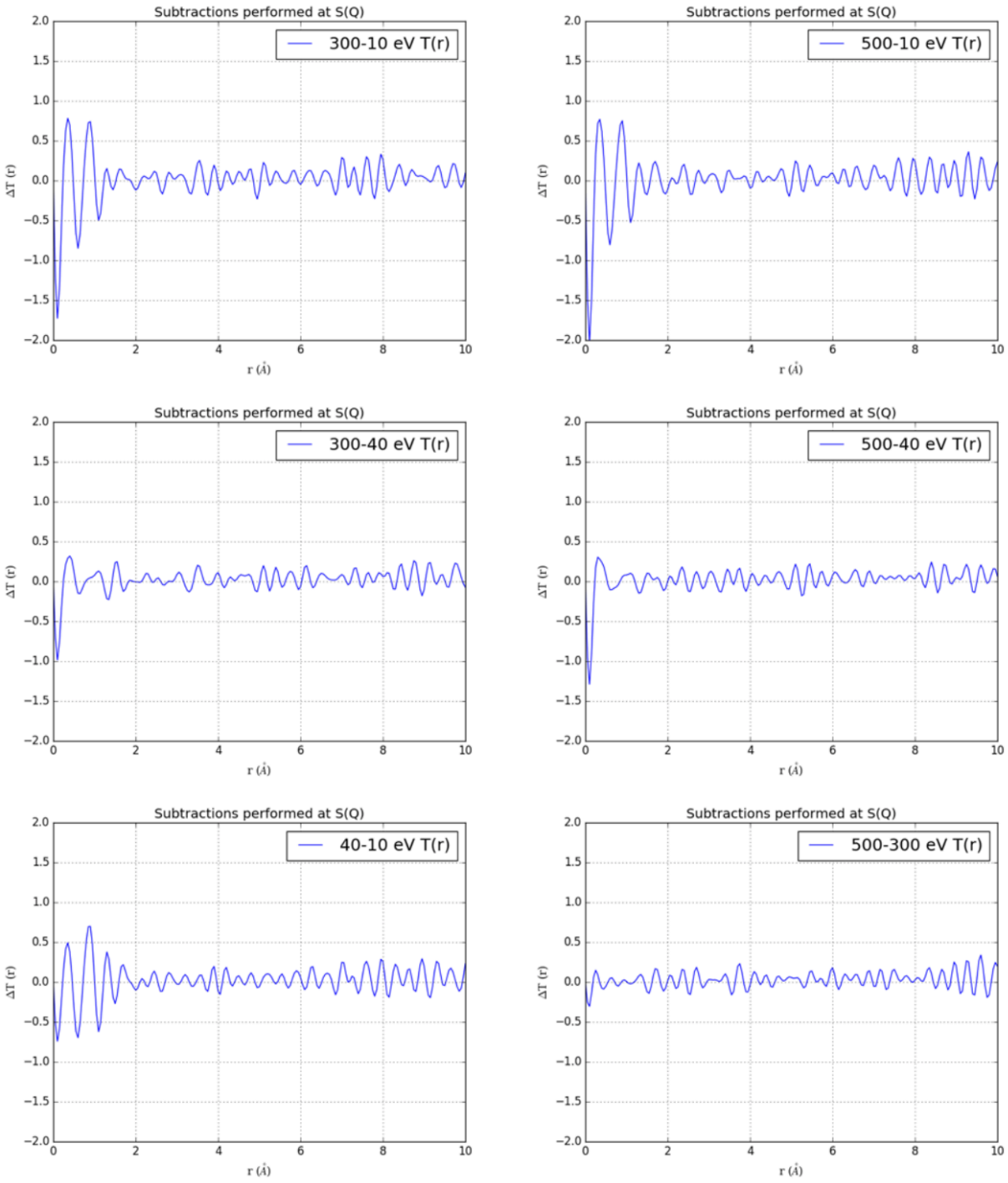


Figure 5.2 *All six possible ΔPDFs created by combining every unique pair of diffraction data sets on* $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$ *collected at X-ray energies of K-10 eV, K-40 eV, K-300 eV, and K-500 eV.*

Statistical correlation indexes were employed to provide a quantitative basis for this observation. To this end, statistical correlations between all possible pairs of ΔPDFs were calculated using three figures of merit, *i.e.*, Pearson, Spearman, and Kendall rank coefficients.[95] Data for these calculations were restricted to the range, r = 2-7 Å. Below this range, only noise is expected, as R···X correlations do not manifest at r = 0-2 Å. At r > 7 Å, the high number of overlapping R···X correlations will render them non-resolvable. Moreover, statistical noise will already be significant in data at r ~ 7 Å or beyond, and both of these problems will only worsen as a function of increasing r. If AXS signals arising from R···X correlations are present in both of the ΔPDF signatures, these established statistical qualifiers should afford positive coefficients between 0 and 1, whereby 1 represents a perfect correlation and 0 should represent the absence of any AXS signal in both data sets. Negative coefficients signify a level of anti-correlation up to a maximum of -1.

The results of these calculations are summarised in Table 5.1. All three sets of coefficients are consistent in their trends, with the Pearson coefficients demonstrating slightly stronger trends than the other two metrics, judging from their slightly higher values overall.

Table 5.1 *Statistical correlations between ΔPDF signatures, calculated in the range r = 2-7 Å, using Pearson, Spearman, and Kendall rank coefficients; these three figures of merit range between 1 and -1 for pure correlation and anti-correlation, respectively (0 represents a null correlation).*

	40–10 eV	300–10 eV	500–10 eV	300–40 eV	500–40 eV	500–300 eV
			Pearson coefficients for 2–7 Å region			
40–10 eV	*n/a*	0.704	0.724	−0.184	−0.221	−0.027
300–10 eV		*n/a*	0.832	0.568	0.299	−0.369
500–10 eV			*n/a*	0.311	0.511	0.208
300–40 eV				*n/a*	0.671	−0.479
500–40 eV					*n/a*	0.328
500–300 eV						*n/a*
			Spearman coefficients for 2–7 Å region			
40–10 eV	*n/a*	0.619	0.618	−0.218	−0.257	−0.047
300–10 eV		*n/a*	0.795	0.582	0.322	−0.362
500–10 eV			*n/a*	0.355	0.544	0.194
300–40 eV				*n/a*	0.671	−0.443
500–40 eV					*n/a*	0.307
500–300 eV						*n/a*
			Kendall rank coefficients for 2–7 Å region			
40–10 eV	*n/a*	0.449	0.446	−0.140	−0.182	−0.040
300–10 eV		*n/a*	0.613	0.410	0.210	−0.259
500–10 eV			*n/a*	0.240	0.370	0.126
300–40 eV				*n/a*	0.483	−0.305
500–40 eV					*n/a*	0.211
500–300 eV						*n/a*

The largest observed positive correlation is between the data sets ΔPDF[(K-500 eV) – (K-10 eV)] and ΔPDF[(K-300 eV) – (K-10 eV)]. This is expected, given that these pairs of ΔPDFs issue the same maximum possible AXS signal of all data sets by virtue of having the K-10 eV data set in common, which possesses the greatest $\Delta f(\lambda)$ value, and the K-300 eV and K-500 eV data counterparts, which are comparable since both of their X-ray energies are sufficiently removed from the Gd K-edge so that $\Delta f(\lambda) = 0$. The statistical correlation that places second highest in rank order of coefficients for each of the three figures of merit is, on average, that pertaining to the relation between ΔPDF[(K-300 eV) – (K-40 eV)] and ΔPDF[(K-500 eV) – (K-40 eV)], *i.e.*, the same comparative data as the highest statistical correlation found, except that the K-40 eV data set replaces the K-10 eV data set in each ΔPDF. This ranking order makes sense once two factors are considered: firstly, while $\Delta f(\lambda)$ will be lower for the K-40 eV data set

than that for K-10 eV data set, it will not be zero, and yet $\Delta f(\lambda)$ will be zero for both the K-300 eV and the K-500 eV data sets; secondly, each ΔPDF component of this correlation will be comparable in signal given that $\Delta f(K{-}300\ eV) = \Delta f(K{-}500\ eV) = 0$, and the strength of the statistical correlations employed herein is based on not only signal intensity between ΔPDFs but also on the relative values of the ΔPDF signals, whereby the more comparable values afford higher statistical correlation. The third- and fourth-highest-ranked correlations relate closely to each other, *i.e.,* they both feature ΔPDF[(K-40 eV) – (K-10 eV)], while its two counterparts are $\Delta PDF[E_{farfromedge} - (K\text{-}10\ eV)]$ where $E_{farfromedge}$ = (K-300 eV) or (K-500 eV), respectively, i.e., those data sets whose X-ray energies are far from the Gd K-edge. The high ranking of these correlations suggests that the K-40 eV data set maps well to its cognate difference pairs K-300 eV and K-500 eV, where $\Delta f(\lambda) = 0$, i.e., that the signal intensities of the ΔPDF pair must be comparable. However, the second-highest-ranked correlation indicated that $\Delta f \neq 0$ for the K-40 eV data set. Taking these two factors into account, the prevailing assumption must then be that $\Delta f(K{-}10\ eV) >> \Delta f(K{-}40\ eV)$ and that $\Delta f(K{-}40\ eV)$ is small but not negligible. The fifth- and sixth-highest-ranked correlations are also related, and assess the level of similarity between $\Delta PDF[E_{farfromedge} - (K\text{-}10\ eV)]$ and $\Delta PDF[E_{farfromedge} - (K\text{-}40\ eV)]$, i.e., the cognate pairs of data sets collected at X-ray energies far from and on the Gd K-edge, where the two energies on the edge are cross-correlated. These are ~ 50% correlated, based on the Pearson and Spearman coefficients, which stands in contrast to the 80% correlation observed between the $\Delta PDF[E_{farfromedge} - (K\text{-}10\ eV)]$ cognate pairs, these fifth and sixth-ranked coefficients effectively measure the cross-correlation between the K-10 eV and K-40 eV. As indicated in discussions above, $\Delta f(K{-}10\ eV) >> \Delta f(K{-}40\ eV)$, i.e., this drop in correlation by 30% reflects this inequality.

There is a distinct drop in coefficient value between this pair of fifth- and sixth-highest-ranked correlations and the seventh- and eighth-highest-ranked correlations. These represent the two-way cross-correlation, comparing $\Delta PDF[(K\text{-}300\ eV) - E_{ontheedge}]$ against $\Delta PDF[(K\text{-}500\ eV) - E_{ontheedge}]$ where $E_{ontheedge}$ for each comparison pair corresponds to a *mix* of the K-10 eV and K-

40 eV data sets. Given that no subtraction components have a single energy in common, their coefficients are both comparable to noise levels (± 30% correlation) as judged by the coefficients that involve ΔPDF[(K-500 eV) – (K-300 eV)], for which Δf = 0. Any ΔPDF correlating against this one will accordingly be comparing against noise (*cf.* the consistently low values of coefficients in the right-hand column of Table 5.1).

In summary, these statistical findings are as expected, i.e., ΔPDF[$E_{farfromedge}$ – (K-10 eV)] signatures exhibit the largest AXS signal, while ΔPDF[$E_{farfromedge}$ – (K-40 eV)] profiles show a modest but significant AXS contribution. All other ΔPDF pairings of data sets acquired at the four X-ray energies manifest negligible AXS signals.

These four ΔPDF data sets that display confirmed AXS signals were taken forward as the essential results of this work, based on which R···X pairwise correlations were determined.

5.6 Scientific Results

Figures 5.3 and 5.4 show overlays of the four ΔPDF signatures that feature confirmed AXS signals, for r = 2-5 Å and r = 4-10 Å, respectively. These data have been partitioned into these two specific regions in order to distinguish structural information that displays R···X correlations, whose average separation, r, is too low for X = R (Figure 5.3), from that that contains R···X correlations, which must involve the R···R separations sought by this AXS study (Figure 5.4).

All four data sets, which contain AXS signals, are presented, rather than just showing the two with largest AXS signal, since all four signals are weak by virtue of being a difference PDF from the diffuse scattering associated with an amorphous material. Moreover, the overarching consideration of the challenging nature of an AXS experiment and its associated data analysis

should be remembered. Accordingly, it was deemed important to display all four sets with an AXS signal, which should present the same AXS signature. Since the results manifest with low signal-to-noise ratio, greater confidence was given to peaks that were present in all four data sets to represent real structural correlations. Peaks from the two strongest AXS signals that overlap with each other (*cf.* those that involve K-10 eV data) but do not correlate with peaks in the cognate K-40 eV pair of data, are nevertheless also considered. Conversely, peak overlap present only in the two AXS signatures that involve K-40 eV data ought not to be considered unless there is independent scientific support for making an exception.

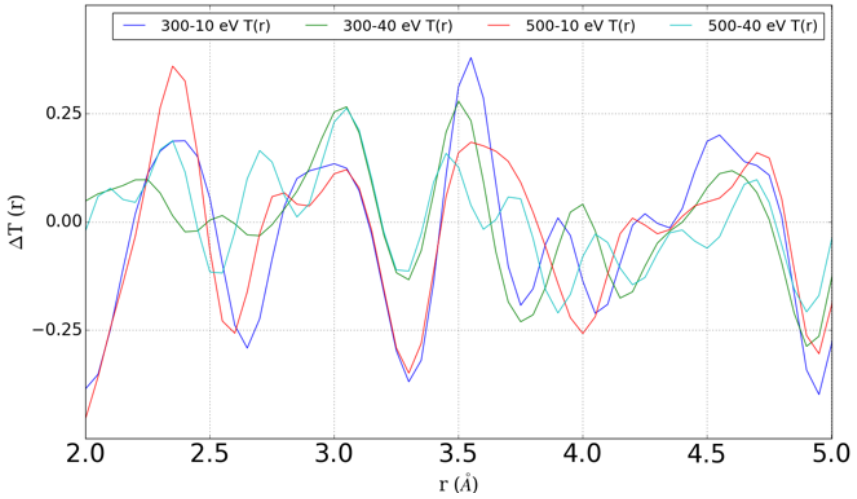


Figure 5.3 *ΔPDFs of the four sets of AXS signals that occur in the range r = 2-5 Å.*

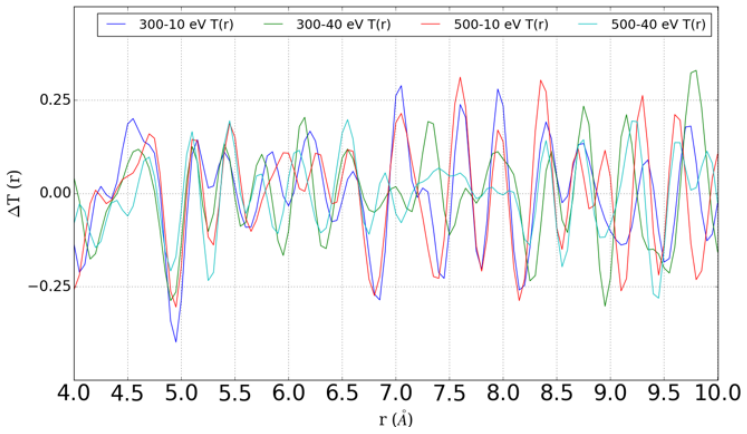


Figure 5.4 *ΔPDFs of the four sets of AXS signals that occur in the range r = 4-10 Å.*

Figure 5.3 should contain R···X (X = P, O) correlations that have already been observed in previous conventional diffraction studies on REPGs,[60-68] where the modelling of atomic pairwise correlations could account for all structural features in the PDF up to r ~ 4 Å. Based on those results, selecting the case where a Gd phosphate glass of a very similar composition (x = 0.229) to the subject sample (x = 0.230) was studied,[62] the following Gd···X correlations should be present in the ΔPDF shown in Figure 5.3: Gd-O at 2.30(2) Å; Gd-(O)-P at 3.64(1) Å. Encouragingly, the ΔPDFs provide evidence for these two pairwise correlations, via the consistent reproduction of peaks at these R···X separations across three or four ΔPDF signatures, respectively. This corroboration between the ΔPDF peak assignments in Figure 5.3 and those from previous conventional diffraction data provide more direct assurance regarding the quality of the AXS data and the analysis methods that afforded them. Moreover, a statistical uncertainty of Δr = 0.1 Å was estimated from the spread of peak values observed between the regions of these ΔPDF profiles that reveal a confirmed pairwise correlation.

Figure 5.4 shows potentially more exciting data since the ΔPDFs therein contain R···X correlations in the region r > 4 Å where we anticipate evidence for R···R separations, the nearest

of which was considered sufficiently important from a materials-centred perspective to motivate this AXS study. Peaks from R···X correlations at 5.1(1) Å and 5.4(1) Å are indicated by the good peak overlap of their AXS signals from all four data sets. Other peaks are suggested at 4.8(1) Å, 6.6(1) Å, 7.1(1) Å, 7.6(1) Å, 7.9(1) Å, 8.4(1) Å and 8.7(1) Å, and their invariance in r with changing Q_{max} (Figure 5.5) suggests that they represent the true local structure.

A weak Gd···Gd pairwise correlation is located at 4.2(1) A, which is representative of a metaphosphate R···R separation. More intense first-neighbour Gd···Gd pairwise correlations are found at the larger radial distributions, 4.8(1), 5.1(1), and 5.4(1) A. These reflect a mixed ultraphosphate and metaphosphate structural character, respectively. These agree with structures of gadolinium ultra-phosphate and gadolinium meta-phosphate, respectively, (*cf.* reference values in Table 5.2).

Table 5.2 *Minimum R···R separations found in crystal structures of rare-earth meta- and ultra-phosphates, references provided in table.*

Rare-earth Meta-phosphate Crystals		**Rare-earth Ultra-phosphate Crystals**	
R (Ref)	**min R···R (Å)**	**R (Ref)**	**min R···R (Å)**
Gd [158]	4.174	Gd [159]	5.171
Gd [160]	5.287	Gd [161]	5.163
		Gd [162]	5.156

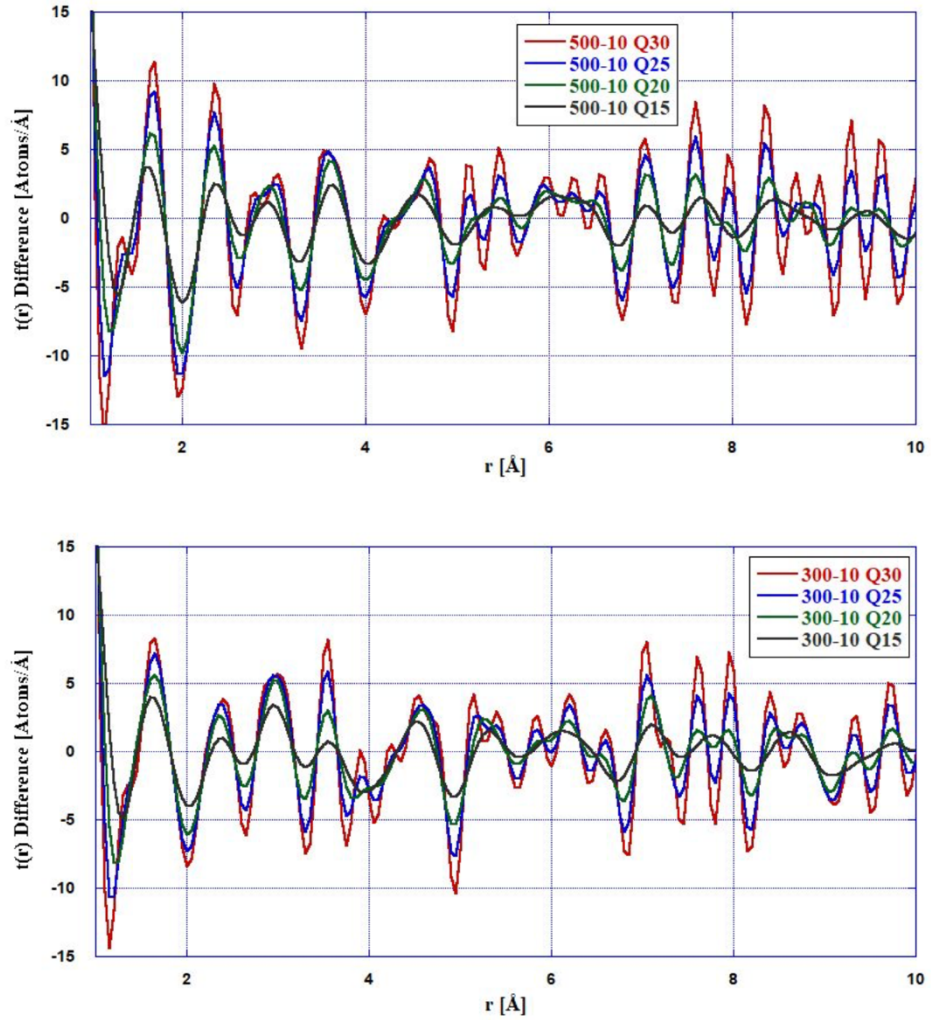


Figure 5.5 *Difference pair distribution functions, ΔPDF [500-10 eV] (top) and ΔPDF [300-10eV] (bottom) with varying Q_{max}: 15 Å^{-1} (black), 20 Å^{-1} (green), 25 Å^{-1} (blue), and 30 Å^{-1} (red). The peak values of r with varying Q_{max} remain unchanged, except at low r, where Fourier series termination effects may persist; this suggests that the features assigned to atomic pairwise correlations in this study are real.*

5.7 Conclusions

The AXS results presented in this study are consistent with findings from other studies and suggest that the examined rare-earth phosphate glasses (REPGs) show mixed ultra- and meta-phosphate structural character when their rare-earth composition lies between the two stoichiometric boundaries of RP_5O_{14} and $R(PO_3)_3$ (x = 0.167 and x = 0.250, respectively, according to the aforementioned formula $(R_2O_3)_x(P_2O_5)_{1-x}$).

The findings of this study have confirmed previous peak assignments and related certain R···R correlations to archetypal meta- or ultra-phosphate structural characteristics: 4.2(1) Å (meta), 4.8(1) Å (ultra/meta), 5.1(1) Å (ultra), 5.4(1) Å (meta), 6.6(1) Å (meta). Moreover, the AXS results have provided the first resolved peak signatures of R···X correlations that lie at average separations of 7.1(1) Å, 7.6(1) Å, 7.9(1) Å, 8.4(1) Å and 8.7(1) Å. These separations could only be classified and quantified by virtue of the high spatial resolution enabled by anomalous X-ray scattering at energies in the region of the Gd K-edge, in contrast to the very broad distributions of REPGs that have previously been observed in the context of two other difference atomic scattering studies,[163,164] both of which use neutrons as the atomic probe. Moreover, the majority of the well-resolved peaks in this study were observed in all four ΔPDFs that each represent the AXS signal, as determined by subtracting X-ray diffraction data collected at an X-ray energy that lies far from the Gd K-edge from a cognate data set acquired at an energy on the edge. Four X-ray energies were employed to obtain the diffraction data for these subtracted pairings, which lie at 10 eV, 40 eV, 300 eV and 500 eV from the Gd K-edge; ΔPDFs involving the K-10 eV data set afforded, as expected, the most intense and reliable AXS signal.

6. Local Structure of SO_2 Photo-isomers in Ruthenium Coordination Complexes

ΔPDF and statistical analysis were performed on data gathered from a synchrotron-based X-ray diffraction experiment on four ruthenium-sulfur dioxide based complexes with a specific interest in light induced structural changes. This chapter is based on a paper being prepared for publication.

My contribution to this project was the analysis and processing of X-ray diffraction data and the interpretation of results, with supervision from my PhD supervisor Jacqueline Cole.

6.1 Summary

SO_2 linkage photo-isomerisation in crystalline ruthenium-based complexes has demonstrated nanophotonic phenomena such as optical switching and nano-optomechanical transduction. Molecular insights into these materials have been explored largely via the characterisation of their photo-induced crystal structures, via *in situ* single-crystal X-ray diffraction, known as photo-crystallography. Photo-induced molecular disorder is present which photo-crystallography can model this to the extent that it is confined within the periodic boundary of a unit cell. However, non-periodic molecular disorder is suspected to exist as well. *In situ* total scattering experiments were therefore carried out, on finely powdered crystals of four ruthenium-sulfur dioxide complexes. Data were modelled using 'light-minus-dark' difference pair distribution function analysis, which afforded photo-induced structural changes exclusively. This revealed structural features that were first compared against models of photo-induced crystal structures known *a priori* from photo-crystallography. Statistical inference was then employed which evidenced generally good agreement between the total scattering data and the photo-crystallographic models, while revealing real differences that are indicative of structure with only short-range order. Overall, our findings demonstrate that *in situ* light-induced

total scattering experiments on finely powdered crystals are able to reveal photo-induced structure. The evidence suggests that such structure could include short-range order as well as photo-crystallographic content. Our demonstration experiment offers a pathway to develop studies that capture short-range order in linkage photo-isomers, while we have outlined the procedure for testing the validity of associated structural models.

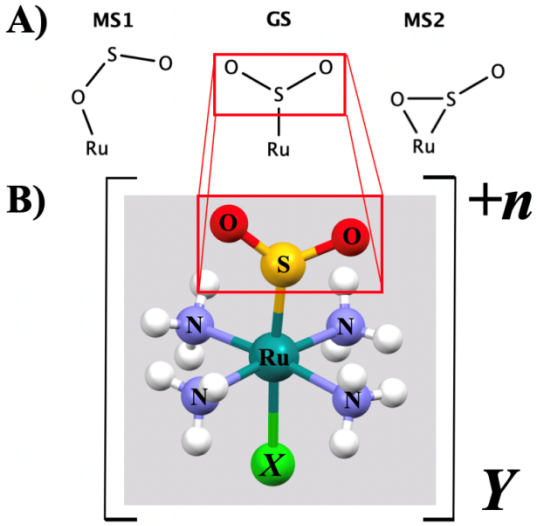
6.2 Introduction

Crystalline materials that exhibit linkage photo-isomerism hold prospects for next-generation photonic and quantum device technologies, given their optical switching[74] and nano-optomechanical transduction[75] characteristics, respectively. These property characteristics are rare in crystalline materials.[165,166] Their origins stem from optically accessible energy levels of linkage photo-isomers in certain coordination complexes that can be induced within a crystal lattice and be maintained in a metastable state once photo-excited. As such, the dark and light-induced states of the photo-active ligand can act as a binary switch: the dark-state ligand configuration signifies "0" and the metastable state isomer signifies "1".[74] Photo-isomerisation within a crystal lattice can have a quite disruptive effect on its immediate crystalline environment, since ligand photo-activation is generally associated with a substantial spatial modulation.[79] Its steric and electronic consequences can be so severe that the photo-activation process triggers a secondary phenomenon within the crystal lattice. For example, the photo-isomerisation of a ligand in the cation of a coordination complex can result in interactions with its neighbouring anion that are so close that the anion moves away in response. Such nano-mechanical transduction has been witnessed in single crystals of a series of ruthenium-sulfur dioxide complexes.[78]

More generally, a wider series of ruthenium-sulfur dioxide complexes have been found to behave as single-crystal optical actuators, exhibiting optical switching[79-85] or nano-optomechanical transduction[78,86] in their single-crystal form. While nanoscopic in its origins, the optical actuation in these complexes often manifests macroscopically in various forms. Its most

commonly observed form is a photo-chromic change in the single crystal, although more dramatic effects such as the light-induced peeling of single crystals have been discovered.[83]

The generic formula of these complexes is $[Ru(SO_2)(NH_3)_4X]Y$, where X is the ligand that lies in the trans position to the photo-active SO_2 ligand. Thereby, up to three distinct SO_2 configurations may co-exist within a crystal structure in different fractions: the η^1-S-bound SO_2 dark state ('GS'), the most thermally stable photo-isomer, an η^2-(O/S)-side-bound SO_2 photo-isomer ('MS2') and the η^1-O-bound SO_2 photo-isomer ('MS1'). Y may be one or more counter-ions, sometimes accompanied by water of crystallisation. (Figure 6.1)



C)

Sample	*X*	*Y*
I	Cl	Cl
II	H_2O	(*p*-tosylate)$_2$
III	4-chloropyridine	Cl_2
IV	3-Cl-pyridine	(*p*-tosylate)$_2$

Figure 6.1 *(A) A schematic display of the SO_2 geometries. The (**MS1**) η^1-O-bound geometry (left), (**GS**) η^1-S-bound geometry (centre), (**MS2**) η^2-S,O-bound geometry (right). (B) Structure of the cation to which the photo-isomerisable SO_2 ligand is attached. H atoms are displayed in white but remain unlabelled for purposes of clarity. (C) A table presenting the exact compositions of the cation and counterion for each sample studied.*

The materials behave as optical switches when they are cryogenically trapped. The dark state represents "0" while either the MS1 or MS2 (or both) photo-isomeric states can typically be generated at 100 K to represent a "1". MS1 only tends to be stable up to about 120 K, whereupon

it converts into the MS2 state. Meanwhile, MS2 is thermally stable to heat up to around 180-220 K where it reverts to the dark state, GS.[80]

Some of the complexes that exhibit MS1 also act as nano-optomechanical transducers at 100 K. This is because of the η^1-O-bound nature of the MS1 photo-isomer: one of its oxygen atoms tends to extrude from its cation, such that it comes unbearably close to some of its neighbouring anions. In certain cases,[78,86] when Y is one or more *para*-methylbenzenesulfonate ('*p*-tosylate') or *para*-chlorobenzenesulfonate anions, the arene ring of one of these counter-ions will rotate about its molecular axis in order to manoeuvre itself away from its neighbouring oxygen atom of the η^1-O-bound SO_2 ligand. This mechanical response alleviates the strain in the arene ring caused by its close proximity to the MS1 SO_2 ligand. The nano-optomechanical transduction process thus manifests as an arene ring rotation that is photo-triggered by the formation of the η^1-O-bound SO_2 ligand.

In situ light-induced single-crystal X-ray diffraction, known as photo-crystallography,[87-90] has been the primary metrological method used to characterise the photo-induced crystal structures of these complexes and to understand the molecular origins of their linkage photo-isomerism. The analysis of photo-crystallographic data uses the well-established strategies for modelling molecular disorder within a periodic boundary of a crystallographic unit cell, to determine the average spatial distributions and photo-conversion fractions of the photo-induced species. Structural perturbations resulting from photo-isomerisation can be determined with sub-angstrom resolution. This has afforded a direct relationship between the photo-induced molecular structures and the corresponding optical properties of these materials.

However, photo-crystallography does not appear to tell the whole story about the structure. Correlated effects between light-induced atomic units are often indicated in photo-crystallographic studies, which cannot properly be modelled as periodic phenomena. Such indications include subtle levels of X-ray diffuse scattering, which typically appears as no more than a small broadening of Bragg peaks; the ostensible splitting of Bragg peaks which can

readily masquerade as twinning; synergic photo-induced mean-square displacements of atoms between (not necessarily bonded) atoms. Such correlated effects are thought to be localised and are expected to show structural signatures at the atomic scale. The short-range order associated with this local atomic structure thus needs probing as well as the crystallographic structure.

The underpinning goal of this study is to probe this structure using pair distribution function analysis of total X-ray scattering data. This method has risen as a technique to solve important problems in nanotechnology, where local atomic structure is key to its understanding.[1] An *in-situ* total X-ray scattering experiment was performed on finely-powdered single crystals of four $[Ru(SO_2)(NH_3)_4X]Y$ complexes, in order to obtain pair distribution functions (PDFs) of their dark- and light-induced structures. More specifically, light-minus-dark difference pair distribution functions were sought since these would yield exclusively the light-induced structural changes in these complexes, while cancelling out various non-structural scattering effects and any systematic errors in the datasets. Difference pair distribution functions are analysed with recourse to two assessments: their structural consistency with photo-crystallography findings so as to explore if pair distribution function analysis could at least reproduce those that stem from this established method; statistical inference of the difference pair distribution functions based on Pearson correlation coefficients to gain insights into purely local atomic structure in these linkage photo-isomeric complexes. The four complexes were chosen as case studies, on the basis of their structural distinctions: (**I**) is the simplest complex in the family, with a single ion in use for X = Y = Cl; it only forms one photo-isomer, MS2, at 100 K; (**II**) is the most stable complex in the series with X = H_2O, Y = *p*-tosylate, whereby Y is not photo-active, and it forms one photo-isomer, MS2, at 100 K. (**III**) contains a substantially larger trans ligand (X = 4-Cl-pyridine) while maintaining small anions; it forms one photo-isomer, MS2, at 100 K. (**IV**) contains the 3-substituted analogue of X in (**III**) which enables it to form predominantly an MS1 SO_2 photo-isomer which triggers nano-optomechanical transduction in its anion, Y = *p*-tosylate, at 100 K. The study will demonstrate that total X-ray scattering experiments on powdered samples of these complexes are able to yield linkage photo-isomerism. We will also show that these results are corroborated by those from photo-crystallography; and

we shall see additional structural features that evidence only short-range order, which may be due to the correlated effects that we seek.

6.3 Methods

All sample preparation and data acquisitions were performed by collaborators.

X-ray diffraction Data Analysis

The 2-D data synchrotron powder diffraction data were integrated and converted to intensity as a function Q using the software, Fit2D [167]. The data were then Fourier transformed to the pair distribution function, *G(r),* via use of local program that applied a Hanning window to taper the intensity function to zero before the Fourier transform in order to minimise the effects of Fourier termination artefacts, with Q_{max} values of: **I**: 19.41, **II**: 17.67, **III**: 14.84 and **IV**: 16.88 Å^{-1}. *G(r)* profiles for sample **IV** were calculated using xPDFSuite [168], since these data did not require use of a Hanning window. This pair distribution function (PDF), *G(r)* relates to the real-space correlation function, ρ^{PDF}, and is defined according to: [93]

$$G(r) = \frac{[\rho^{PDF}(r) - \rho_0]}{\rho_0}\left(\sum_{i=1}^{n} c_i \bar{f}_i\right)^2 \quad (6.1)$$

where ρ_0 is the average number density of the subject material (in Atoms Å^{-3}), c is the composition of species, i, in the material, *f* is the X-ray atomic scattering factor, and $\rho^{PDF}(r)$ is a real-space correlation function, such that:

$$\rho^{PDF}(r) - \rho_0 = \frac{1}{2\pi^2 r}\int_0^{\infty} Q[S(Q) - 1]sinQrdQ \quad (6.2)$$

where $Q = (4\pi \sin\theta)/\lambda$ is the X-ray scattering vector that affords the structure factor, S(Q), that has an associated atomic distance, r, in real space.

Data Model Simulations

Simulating G(r) Models

Simulated models of *G(r)* were created for each compound before and after photo-irradiation for the purpose of comparison with the experimental findings.

Model structures of each sample were created using their dark and light-state atomic coordinates [80,81,86] that have been determined previously via photo-crystallography.[89,90] Their crystal structures afforded from photo-crystallography were refined within the periodic boundary of their respective unit cell; as such, their co-existing dark and light components were modelled as molecular disorder, using a fractional occupancy parameter to refine the population of each light-induced species and the residual dark state. Thus, the SO_2 ligand in samples **I-III** was modelled as disorder with its dark state ("GS") and one light-induced state ("MS2") atomic coordinates being refined with a population, x% GS and (1-x)% MS2. Sample **IV** exhibits two light-induced states, "MS1" as well as MS2; thus, its SO_2 ligand is refined with three sets of atomic coordinates and occupancy factors, GS, MS1 and MS2, whose overall population sums to 100%. Light-induced molecular transduction also occurs in sample **IV**, whereby one of its tosylate counterions is also disordered in its MS1 state. The molecular transduction phenomenon is triggered by the formation of MS1 SO_2 species. Yet, the disorder in the affected tosylate ion is refined independently from the SO_2 disorder model, since its onset and delay is not proportional to that of MS1.[78] As such, the population of this tosylate ion in the light-state y% is balanced by its (1-y)% dark-state structure.

The first step in the construction of these new models was separating the atomic coordinates of each of the constituent parts of this molecular disorder model for GS, MS2 (and

where present, MS1) and appending these to the atomic coordinates of the rest of the coordination complex to generate molecular structures of each sample in each SO_2 configuration. The molecular structures of the MS1 and MS2 states were then manually expanded into supercells (4 x 4) using VESTA 3 [169]. This manual intervention was required since the molecular disorder of the fractional light-induced states prevented xPDFSuite [168] from automatically using the crystallographic symmetry of the structures to afford the task at hand. The use of supercells ensured that inter-ionic interactions could also be incorporated into each light-induced structural model. The dark-state models did not feature molecular disorder or fractional occupancies and so they did not require manual expansion into supercells. The resulting structures were used as the inputs for xPDFSuite [168], to simulate $G_{sim}(r)$ profiles, without modification by any refinement with respect to lattice parameters, atomic displacement parameters, or correction due to instrumental artefacts, such as peak dampening or broadening. This afforded simulated *G(r)* profiles in the form of $G_{sim}(r)_{\mathrm{GS}}$ and $G_{sim}(r)_{\mathrm{MS2}}$ for each sample in each state, and additionally $G_{sim}(r)_{\mathrm{MS1}}$ for the MS1 state of sample **IV**.

Simulating $\Delta G_{sim}(r)_{light-dark}$ Models

The difference pair distribution function, $\Delta G_{sim}(r)_{light-dark}$, was also calculated by subtracting the dark-state *G(r)* contribution from that of the light-induced state. In theory, all of the signal in this difference should represent changes in local atomic structure that occur exclusively as a result of photo-irradiation; as such, the difference could be calculated according to:

$$\Delta G_{sim}(r)_{light} = A[G_{sim}(r)_{MS2}] + B[G_{sim}(r)_{MS1}] + [1 - (A + B)]\,[G_{sim}(r)_{GS}] \quad (6.3)$$

$$G_{sim}(r)_{dark} = G_{sim}(r)_{GS} \quad (6.4)$$

$$\Delta G_{sim}(r)_{light-dark} = A\,[G_{sim}(r)_{MS2} - G_{sim}(r)_{GS}] + B\,[G_{sim}(r)_{MS1} - G_{sim}(r)_{GS}] \quad (6.5)$$

For samples **I-III**, an MS1 state does not exist, so B = 0 and these equations simplify to:

$$\Delta G_{sim}(r)_{light} = A[G_{sim}(r)_{MS2}] + (1 - A)[G_{sim}(r)_{GS}] \quad (6.6)$$

$$G_{sim}(r)_{dark} = G_{sim}(r)_{GS} \quad (6.7)$$

$$\Delta G_{sim}(r)_{light-dark} = A\,[G_{sim}(r)_{MS2} - G_{sim}(r)_{GS}] \quad (6.8)$$

Within the scope of this theoretical construct, these coefficients A and B could be thought of as population parameters for each dark or light-state configuration of a sample that could be refined against the experimental data to afford a structural model. However, this construct assumes that contributions of dark and light-induced states to the scattering intensity are kept equal and constant throughout data collection. This would require constant beam flux as well as strict control of factors such as exposure time, detector response time and background artefacts. In other words, one would have to be sure that any change in the number of counts reported by the detector before and after sample irradiation is due to a physical change in the material and is unaffected by experimental artefacts. Otherwise, light-induced state structure is contaminated by some residual scattering intensity in its $\Delta G_{sim}(r)_{light-dark}$ profile.

It was not possible to satisfy such a mandate for this experiment, in part because this study represents the first reported attempt of a PDF-based photo-crystallographic experiment which was already challenging in its own right. Nonetheless, $\Delta G_{sim}(r)_{light-dark}$ could be calculated using a similar form of equations, which correct for these experimental artefacts by normalizing the contributions of each dark and light-induced state, using coefficients:

$$\Delta G_{sim}(r)_{light-dark} = C_1\,[C_2(G_{sim}(r)_{MS2}) - G_{sim}(r)_{GS}] \quad (6.9)$$

where only GS and MS2 states are present, as is the case for samples **I-III**, while:

$$\Delta G_{sim}(r)_{light-dark} = C_1[C_2(G_{sim}(r)_{MS2}) + C_3(G_{sim}(r)_{MS1}) - G_{sim}(r)_{GS}] \quad (6.10)$$

for cases such as sample **IV**, where GS, MS2 and MS1 states are present.

The coefficients $C_1,\ C_2,\ C_3$ were calculated by maximizing the Pearson correlation coefficients [95] between $\Delta G_{sim}(r)_{light-dark}$ and $\Delta G_{exp}(r)_{light-dark}$ using the generalised reduced gradient method of Microsoft Excel Solver [170], over a low-r region (Region 1 in Table 2) which describes the bond lengths most closely associated with the SO_2 ligand. Pearson correlation coefficients, ρ, are statistical measures that have previously been used to quantify the agreement between difference pair distribution functions [98-100]. ρ can range between -1 and 1, the former signifying a perfectly negative correlation and the latter a perfectly positive correlation. A coefficient of 0 signifies no correlation at all. [95]

The construction of $\Delta G_{sim}(r)_{light-dark}$ in this manner, using independent structural configuration models of pure GS, MS1 or MS2 states, is analogous to having a completely heterogeneous partially photo-isomerised sample. This model reproduces atomic pairwise correlations within the ionic structures, as well as inter-ionic interactions that feature 'like-for-like' SO_2 configurations in neighbouring cations (MS2···MS2, MS1···MS1, GS···GS). However, it does not account for inter-ionic interactions between cations whose neighbours have 'mixed' SO_2 configurations, *i.e*. pairwise correlations that relate to different SO_2 configurations that are neighbours of each other such that MS2···GS, MS1···GS or MS1···MS2 inter-ionic interactions form. However, since the spatial distribution of photo-isomerised molecules throughout a sample is unknown, it is not possible to construct a model that incorporates these inter-ionic interactions from complexes with disparate SO_2 configurations. The effects of inter-ionic interactions on the $\Delta G(r)$ are discussed in the Results section.

6.4 Results and Discussion

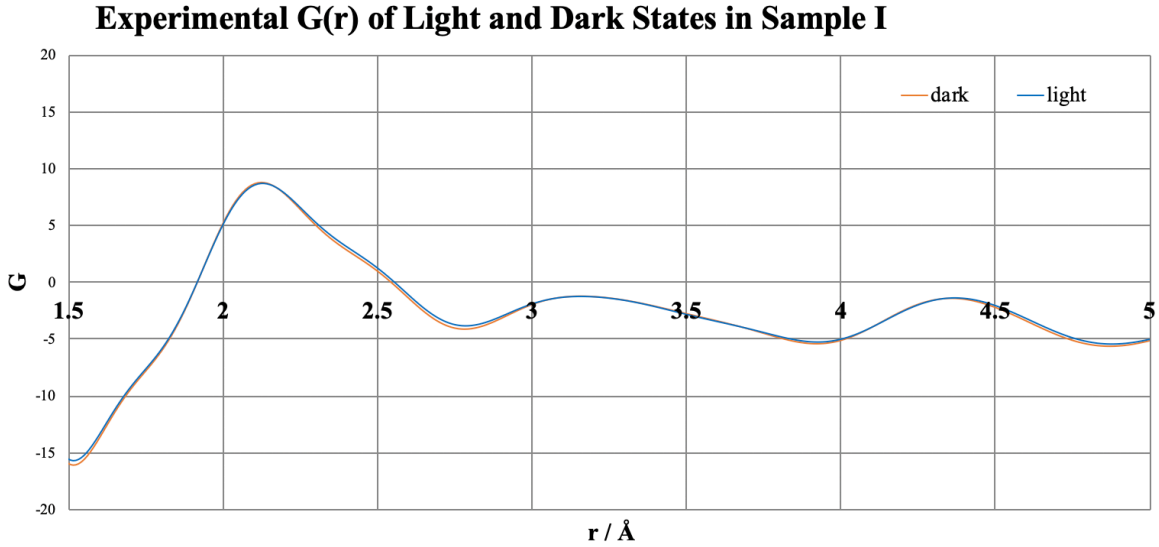
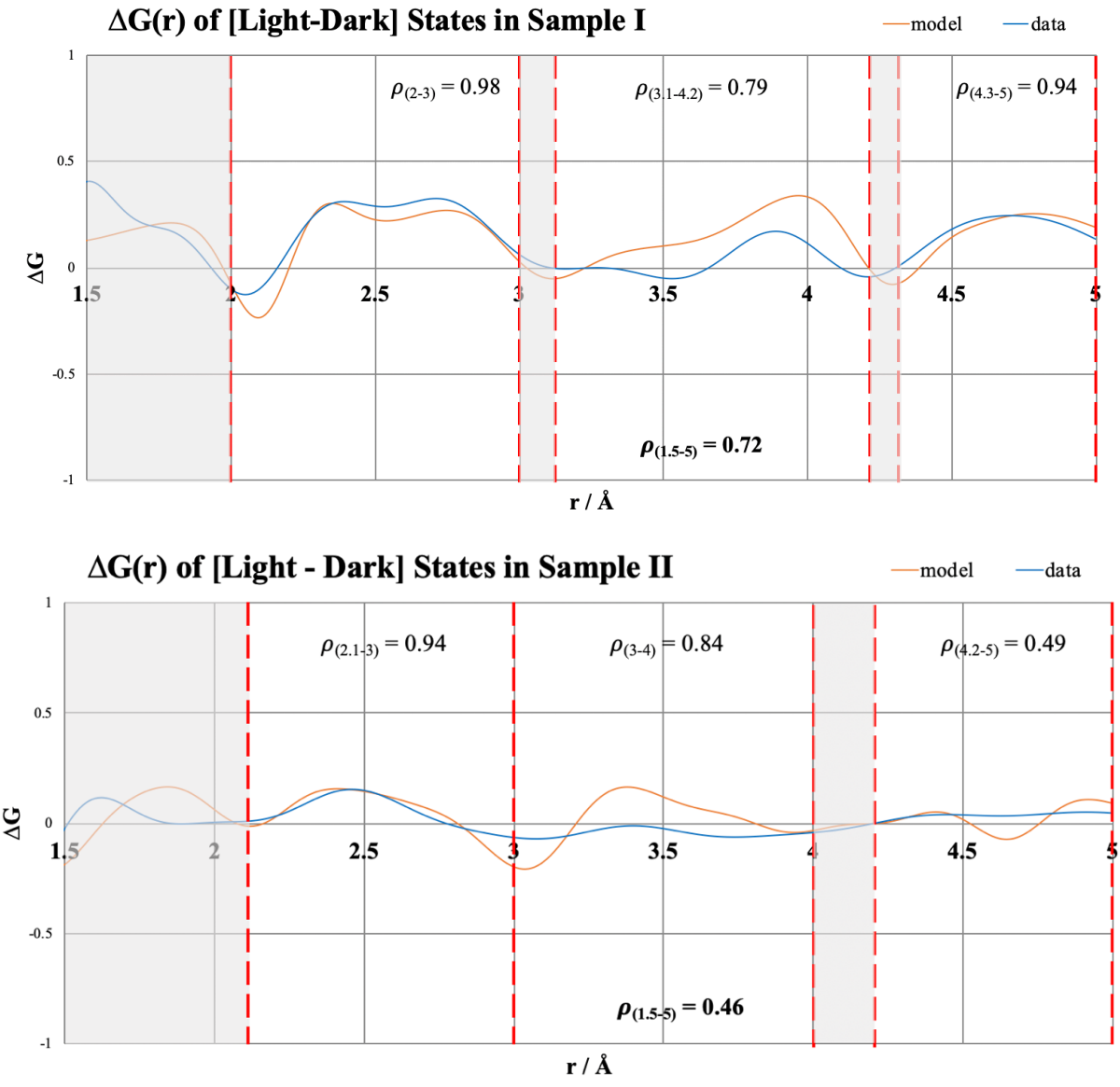


Figure 6.2 *$G_{exp}(r)_{light}$ (orange) and $G_{exp}(r)_{dark}$ (blue) for sample* ***I.***

As can be seen in Figure 6.2, the $G_{exp}(r)$ profiles for the dark and light-induced state of sample **I** look very similar, with a Pearson correlation coefficient of ρ = 0.99964 over the range r = [1.5 – 5] Å, where a value of 1 would indicate a perfect correlation. Analogous *G(r)* profiles for the other three samples showed similarly small differences; their profiles are given in Appendix A2. Given these very subtle differences, and the intrinsic nature of this experiment which seeks a difference between light and dark-state structure, data analysis via the use of difference pair distribution functions presents an attractive option. Moreover, $\Delta G(r)_{light-dark}$ profiles will only afford atomic pairwise correlations that are due to the light-induced structural changes. So, this substantially simplifies what would otherwise be a set of *G(r)* profiles that display heavily overlapping atomic pairwise correlations from all possible structural detail in these ruthenium-based complexes. An analysis of $\Delta G(r)_{light-dark}$ rather than *G(r)* profiles also circumvents the need to correct the data for effects such as X-ray absorption or Compton scattering or systematic bias, which are likely to be present judging from the slight off-sets of

these curves. These effects can be considered to be equally present in the light-induced and dark-state data of these samples within the margin of experimental error, and will thus cancel upon producing the difference pair distribution function.

Accordingly, $\Delta G(r)_{light-dark}$ profiles were determined for samples **I-IV**, as shown in Figure 6.3.



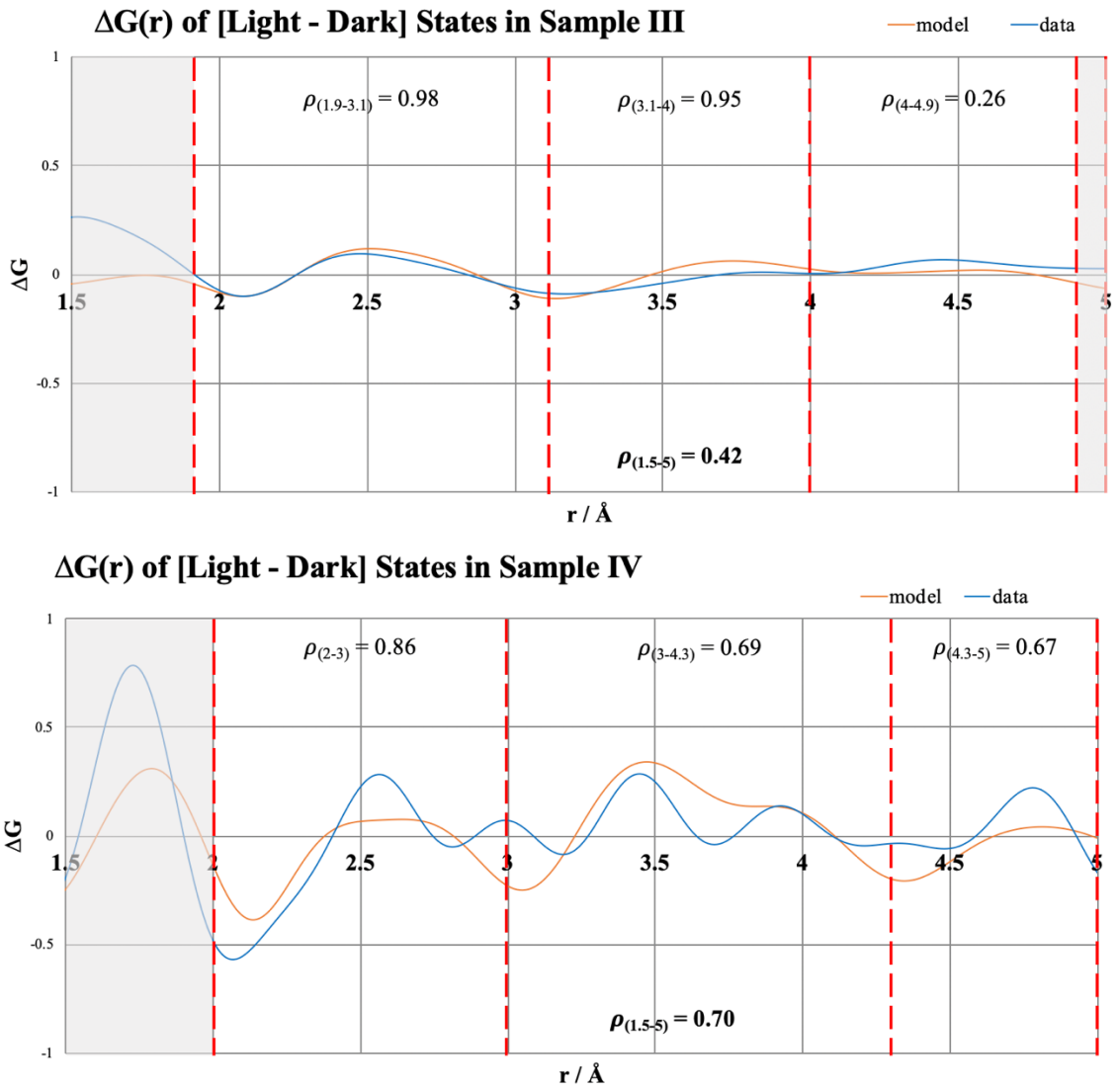


Figure 6.3 *$\Delta G(r)_{light-dark}$ for samples **I-IV** generated from experimental data (blue) and from simulated models (orange) that account for the local structural changes accompanying photo-isomerisation that are also present in photo-crystallographic models. The r-range below 1.5 Å is omitted since it predominantly features artefacts arising from Fourier transform series termination effects.*

Prior to any quantitative analysis, it is worth noting that the pure existence of structural features in the experimental $\Delta G_{exp}(r)_{light-dark}$ (blue profiles in Figure 6.3) can be used to evidence the existence of local atomic structure in these samples as well as demonstrate that photo-isomerisation is achievable in these compounds in their polycrystalline form; even when single-crystal samples are crushed into powders and measured in a spinning capillary setup.

Quantitative analysis of the $\Delta G_{exp}(r)_{light-dark}$ profiles can be considered in terms of two types of structural changes: those that display *both* short- and long-range order; and those that exhibit *only* local atomic structure. The former type is expected to arise from the well-known light-induced molecular disorder that causes bond-geometry changes between dark- and light-induced crystal structures of this series of ruthenium-sulfur dioxide complexes**,** as determined via photo-crystallography. The dark and light-induced crystal structures of **I-IV** thus formed the basis of construction for the simulated models, $\Delta G_{sim}(r)_{light-dark}$, in this study. The regions of the crystal structures of samples **I-IV** which are most affected by the photo-isomerisation are particularly helpful in identifying first (direct) and second-neighbour atomic pairwise correlations that are associated with this molecular disorder. These regions essentially envelop the immediate ruthenium core environment for samples **I-III;** the same situation befits sample **IV,** except that one of the *p*-tosylate counterions of **IV** is also affected substantially by photo-isomerisation. Figures 6.4 and 6.5 depict the first (direct) and second-neighbour atomic pairwise corrections that are expected to arise in these ruthenium core environments of samples **I-IV**, based on constructions from the photo-crystallography results.

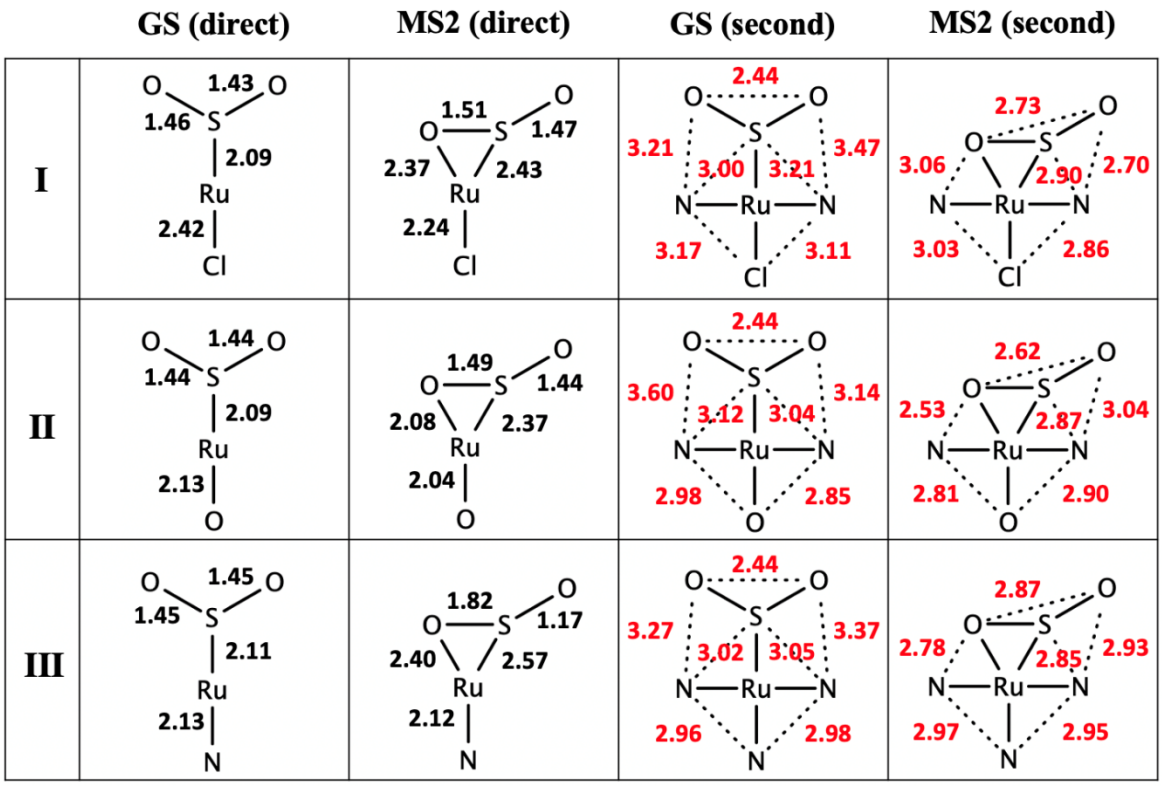


Figure 6.4. *Direct (black) and second-neighbour (red) atomic pairwise separations (Å) of the immediate ruthenium core environment of samples* ***I-III****, in their dark (GS) and light-induced (MS2) state configurations; this core is the structural region of* ***I-III*** *that is most affected by photo-isomerisation, judging from photo-crystallographic results.* [80,81] *Atomic pairwise correlations that are less affected by photo-isomerisation are omitted for purposes of clarity. Presented bond geometries were taken from light-induced single-crystal structures of* ***I:*** [80]***, II:*** [80]***, III:*** [81]***,*** *determined by photo-crystallography.*

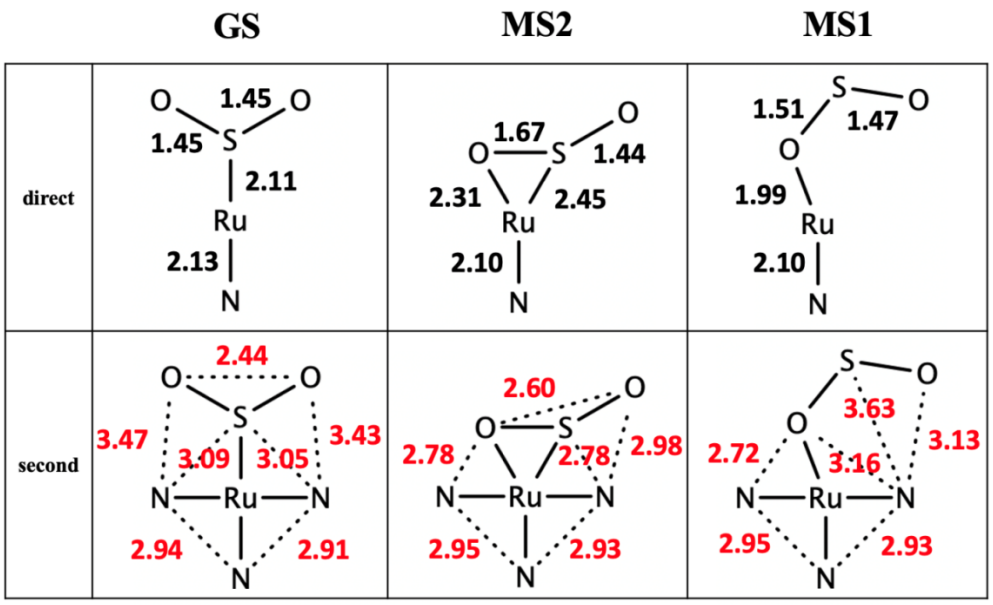


Figure 6.5 *Direct (black) and second-neighbour (red) atomic pairwise separations (Å) of the immediate ruthenium core environment of sample* ***IV****, in its dark (GS) and light-induced (MS2 and MS1) state configurations; this core is the structural region of* ***IV*** *that is most affected by photo-isomerisation, judging from photo-crystallographic results.* [86] *Atomic pairwise correlations that are less affected by photo-isomerisation are omitted for purposes of clarity. Presented bond geometries were taken from the light-induced single-crystal structure of* ***IV:*** [86]**,** *determined by photo-crystallography.*

Figures 6.4 and 6.5 show a large number of atomic separations of similar length. This number is sufficiently large to render unfeasible the assignment of individual peaks in the $\Delta G_{exp}(r)_{light-dark}$ profiles to specific pairwise corrections. Moreover, some of the atomic separations in Samples **I-IV** contract while others expand upon photo-irradiation. Thus, light-induced structural changes will manifest as a complicated set of negative and positive features in a $\Delta G_{exp}(r)_{light-dark}$ profile, some of which may partially or wholly cancel another. This makes the ability to even observe individual peak contributions challenging qualitatively, let alone unravel quantitative information about any atomic pairwise correlation. Nonetheless, the light-induced changes responsible for certain low-r features in these profiles can still be rationalised on the basis of the relative level of consistency between atomic pairwise correlations shown in

the $\Delta G_{exp}(r)_{light-dark}$ profiles and cognate light-induced structural changes deduced previously via photo-crystallography; or on the basis of various common structural features that co-exist among the samples.

Structural consistency between $\Delta G_{exp}(r)_{light-dark}$ and photo-crystallography findings

The $\Delta G_{exp}(r)_{light-dark}$ of sample **I** displays a negative peak centred at 2.05 Å and a positive peak centred at 2.39 Å. These peaks are consistent with photo-crystallography results, whereby light-induced depletion of the 2.09 Å Ru–S bond (GS), yields two 2.37 Å Ru–O and 2.43 Å Ru–S bonds (MS2) and a concomitant shortening of the Ru–Cl bond from 2.42 Å (GS) to 2.24 Å (MS2). The positive peak centred at 2.71 Å and negative peak centred at 3.16 Å could be attributed to the respective emergence and depletion of structural features due to photo-irradiation that result in an overall shortening of second-neighbour atomic pairwise correlations; *cf.* on average, the value of such correlations for sample **I** is 2.88 Å for MS2 and 3.09 Å for GS.

These low-r structural features in $\Delta G(r)_{light-dark}$ are broadly similar across samples **I-III,** and are consistent with their cognate photo-crystallographic findings, as shown in Figures 6.3 and 6.4 respectively. Therefore, it can be surmised that samples **II** and **III** exhibit analogous structural changes. However, subtle differences do exist.

For example, the $\Delta G_{exp}(r)_{light-dark}$ profile of sample **II** within the range [2-2.5] Å is similar to that of sample **I**, albeit with a smaller amplitude and greater breadth. This stands to reason since the 2.09 Å Ru–S bond (GS) in sample **II** gives way to 2.08 Å Ru–O bonds (as opposed to 2.37 Å Ru-O bonds in sample **I**) and 2.37 Å Ru–S bonds (MS2) upon photo-irradiation, resulting in an analogous, but wider peak profile of a lesser magnitude.

Over the [2.5-3] Å range, the $\Delta G_{exp}(r)_{light-dark}$ profile of sample **II** ostensibly appears quite different to that of sample **I**. However, the averages of its second neighbour atomic

pairwise correlations shown in Figure 6.4 are 3.02 Å for GS and 2.80 Å for MS2, which is essentially in line with those of sample **I**. These values are slightly shorter for sample **II** than sample **I**, which is presumably a simple consequence of the smaller *trans* influence in sample **II**; *cf.* the ligand *trans* to SO_2 is Cl and OH_2 in samples **I** and **II**, respectively. The visual discrepancy observed between sample **I** and **II** is thus the result of a distinct left-shifting of the associated negative and positive peaks that arise from these two respective correlations. Correspondingly, the $\Delta G_{exp}(r)_{light-dark}$ profile of sample **II** in this range of r drops off as a function of increasing r much sooner than it does in sample **I**, while the compensatory negative peak that emerges in sample **II**, centred at 3.06 Å, follows suit with sample **I**.

The $\Delta G_{exp}(r)_{light-dark}$ profile of sample **III** is consistent with the photo-crystallographic observation that a 2.11 Å Ru–S bond (GS) gives way to 2.40 Å Ru–O and 2.57 Å Ru–S bonds (MS2) upon photo-irradiation. This structural change is also analogous to that observed for samples **I** and **II**, except that the Ru-O and Ru-S bonds in the MS2 configuration of sample **III** are distinctly longer than those in samples **I** and **II**. The somewhat different (N-substituted 4-chloropyridine) *trans*-ligand to SO_2 for sample **III** is presumably a significant cause of its weaker SO_2 linkage photo-isomer.[81] The $\Delta G_{exp}(r)_{light-dark}$ profile of sample **III** in the 2.0-2.5 Å region correspondingly exhibits a negative peak at 2.08 Å, and a positive peak at 2.48 Å that is distinctly right-shifted relative to those of samples **I** and **II**. Over the [2.5-3] Å range, the $\Delta G(r)$ profile of sample **III** is quite similar to that of sample **II**, evidencing a broad peak that drops off as a function of increasing r until its compensatory negative peak emerges as the dominant feature, centred at 3.17 Å.

The $\Delta G(r)_{light-dark}$ profile of sample **IV** appears somewhat contrary, when compared against those of samples **I-III**. Its $\Delta G_{exp}(r)_{light-dark}$ similarly dips at 2.06 Å, but to a far greater extent than seen in samples I-III; its negative peak also displays a greater span, reaching a distinctly higher value of r before $\Delta G_{exp}(r)_{light-dark}$ becomes positive at 2.45 Å. This net positive $\Delta G_{exp}(r)_{light-dark}$ contribution does not peak until 2.56 Å, which is significantly

higher in r, compared with the $\Delta G(r)_{light-dark}$ profiles of samples **I-III**. This contrary finding stands to reason given that the linkage photo-isomerisation results are markedly different in sample **IV**. In particular, the MS1 SO_2 photo-isomer predominates in the light-induced form of sample **IV**, judging from the photoconversion fractions that were observed from photo-crystallography, *cf.* 36.2% MS1, 6.9% MS2, 56.9% GS at 100 K.[78] The MS2 SO_2 configuration seen in samples **I-III** is also present in sample IV, but it appears in a modest quantity. The $\Delta G_{exp}(r)_{light-dark}$ profile of sample **IV** corroborates the photo-crystallography findings by manifesting a similar predominance of MS1 photo-isomerisation: the marked dip in $\Delta G(r)$ at 2.06 Å is consistent with a light-dark bond contraction that results from a depletion of the 2.11 Å Ru-S bond (GS) and the concurrent emergence of a 1.99 Å Ru-O bond (MS1). The nature of this dip suggests that the MS1 SO_2 configuration forms with a similarly high photoconversion fraction to that which was observed by photo-crystallography. [78] The observation that $\Delta G_{exp}(r)_{light-dark}$ only becomes net positive by 2.45 A, and does not peak until 2.56 Å, is symptomatic of a predominantly MS1 SO2 photo-isomerisation. Notwithstanding minor contributions from MS2 bond geometry that might arise, Figure 6.5 shows that there are no direct bonds in the immediate ruthenium core environment of sample **IV** above 2.13 Å which are significantly affected by photo-isomerisation; meanwhile, any atomic pairwise correlation at or below 2.13 Å will be masked by the aforementioned $\Delta G(r)$ contributions of the 2.11 Å Ru-S bond and the 1.99 Å Ru-O bond (MS1). The shortest second-neighbour atomic pairwise correlation, from the GS and MS1 SO_2 configurations of sample **IV** shown in Figure 6.5, will present a positive $\Delta G(r)$ contribution is centred at 2.72 Å. This value of r is precisely where the $\Delta G_{exp}(r)_{light-dark}$ profile finally shows a positive peak within its 2-3 Å region.

The next part of the $\Delta G(r)$ profile is also different to samples **I-III** since the structural features of the average second-neighbour atomic pairwise correlations which manifest in samples **I-III** in the region of about 2.5-3.1 Å differ owing to the presence of the MS1 SO_2 configuration.

All four samples in this study will exhibit third-neighbour intra-ionic pairwise correlations as well as inter-ionic interactions above ~ 3 Å. Thus, any further analysis of this comparative type between $\Delta G_{exp}(r)_{light-dark}$ and photo-crystallography findings was precluded beyond this value of r.

In any case, this type of comparative analysis against photo-crystallographic models can only serve the data analysis so far, since it justifies $\Delta G_{exp}(r)_{light-dark}$ based on light-induced molecular disorder that exhibits *both* short and long-range order. These comparisons are important since they provide good consistency between the results for samples **I-IV**. They thus provide assurance that the experimental data are legitimate. They also confirm that photo-isomerisation can indeed take place in a polycrystalline powder, and do so in a fashion that is similar to that which one expects in photo-crystallography experiments. However, it has yet to be assessed if there are atomic pairwise correlations in samples **I-IV** which exhibit *only* local structure order. One way to gain some insights towards such an assessment is to analyse the level of statistical agreement between the experimental data and the simulated models of $\Delta G(r)$. This works because the models were drawn from the photo-crystallography results, while the data herein may yield not only short- and long-range order that is characteristic of photo-crystallography findings; it may also potentially exhibit $\Delta G_{exp}(r)_{light-dark}$ contributions that are due solely to short-range order.

A form of statistical inference which correlates the level of agreement between the $\Delta G_{sim}(r)$ and $\Delta G_{exp}(r)$ is now employed to analyse the $\Delta G(r)$ profiles of samples **I-IV**, across a wider range of r than the ~ 3 Å limit mentioned in the approach described above. Moreover, it is demonstrated that this statistical inference method can identify differences between $\Delta G_{exp}(r)_{light-dark}$ and $\Delta G_{sim}(r)_{light-dark}$ that can provide insights into local atomic structure that may *only* exhibit short-range order.

Statistical inference of Δ*G(r)* profiles based on Pearson correlation coefficients

This analytical method is performed by dividing the Δ*G(r)* into three regions. The exact range of each region is sample dependent, with the first (**~ [2-3] Å)** encompassing direct bonds that are associated with the SO_2 and its *trans*-ligand, where the effects of photo-isomerisation are most evident. The second region (~ **[3-4] Å**) incorporates second-neighbour atomic pairwise correlations from intra-ionic interactions, some higher-order neighbour intra-ionic correlations, as well as correlations from close inter-ionic interactions. The third **(~ [4-5] Å)** envelops the longest intra-ionic as well as most remaining inter-ionic atomic pairwise correlations. The ranges are set with the intention of preventing the splitting of peaks across separate regions, as far as this is possible.

Table 2. *Exact ranges of sample-dependent Δ*G(r)* subdivisions.*

Sample	**Region 1 (Å)**	**Region 2 (Å)**	**Region 3 (Å)**
I	[2 - 3]	[3.1 - 4.2]	[4.3 - 5]
II	[2.1 - 3]	[3 - 4]	[4.2 - 5]
III	[1.9 - 3.1]	[3.1 - 4]	[4 - 4.9]
IV	[2 - 3]	[3 - 4.3]	[4.3 -5]

Dividing Δ*G(r)* in this manner facilitates the identification of the complex structural changes associated with photo-isomerisation, as well as enabling a validation of the constructed models. The validity each simulated model was quantified by calculating Pearson correlation coefficients over each region. The resulting Pearson coefficients can be seen in Table 2 and as annotations in Figure 6.3.

Table 3. *Pearson correlation coefficients (ρ) between experimental data and simulated models of ΔG(r) presented over three regions of interest as well as providing the overall value.*

Sample	ρ_1	ρ_2	ρ_3	ρ_{ovr}
I	0.98	0.79	0.94	0.72
II	0.94	0.84	**0.49**	0.46
III	0.98	0.95	**0.26**	0.42
IV	0.86	**0.69**	**0.67**	0.70

As can be seen from the values in Table 2, good agreement can be observed in the two low-r regions (~ **[2-3] Å** and ~ **[3-4] Å,** $\boldsymbol{\rho}_1$, $\boldsymbol{\rho}_2$ in Table 2) for all samples, except for $\boldsymbol{\rho}_2$ of sample **IV**. These are the two regions where the bonds most representative of the structural change of photo-isomerisation are expected to manifest themselves (Table 3). The physical significances of the features in Region 1 were described earlier in this chapter.

The low ρ values in Region 3 may well be due to the greater difficulty in modelling data in this higher range of r where the photo-induced structural changes of many pairwise correlations will manifest. However, there is another possible interpretation to consider, particularly since the ρ value in Region 2 of sample **IV** is also damped, as we now discuss.

Regions 2 and, especially, 3 are expected to be influenced by atomic pairwise correlations that stem from inter-ionic interactions. As previously mentioned, the simulated $\Delta G_{sim}(r)$ profiles were created using models of supercells, featuring either solely photo-isomerised (MS1/MS2) or non-photo-isomerised (GS) molecules. This unavoidable limit in the models led to their absence of most types of 'mixed' inter-ionic interactions that are expected to exist in the samples; the models accounted only for like-like type (GS-GS, MS1-MS1 or MS2-MS2) inter-ionic interactions. The impact of this issue on ρ, arising from a lack of agreement

between experimental data and models, would be minimised if photo-isomerisation were to occur in the samples in large domains. This is because most inter-ionic interactions would be of the 'like-like' type within the large domain of a crystal structure whilst 'mixed' types of interactions (GS-MS1, GS-MS2, MS1-MS2) would exist only on the periphery of these domains, thus being minimal in quantity and barely measurable. However, the low ρ values observed suggest that this may not be the case; rather, domains could be sufficiently small to allow 'mixed' type inter-ionic interactions to have a significant effect on $\Delta G_{exp}(r)$.

The markedly poorer agreement between experimental data and the models for sample **IV,** as manifested by low ρ values in *both* Regions 2 and 3, somewhat substantiates the interpretation of an influence of structure with short-range order arising from 'mixed' types of inter-ionic interactions. Such an influence would be especially prominent in sample **IV**, since it is the only sample that exhibits both the MS1 and MS2 light-induced states, thus featuring all three 'mixed' types of inter-ionic interactions. Moreover, the MS1 state is by far the most populous, judging from the 36.2 : 6.9 MS1 : MS2 photo-conversion fractions observed by photo-crystallography.

Additionally, the formation of the MS1 state triggers a phenomenon referred to as a light-induced nano-optomechanical transduction, which leads to a complementary rotation in one of the *p*-tosylate counterions. This introduces an entirely new set of light-induced structural changes at the low-r region, in addition to those associated with the SO_2 ligand. These additional changes are particularly difficult to analyse because they are primarily inter-ionic, between an SO_2 ligand in its MS1 state and its closest neighbouring *p*-tosylate counterion. These inter-ionic interactions are especially hard to model in such compounds, for reasons discussed above. Additionally, the disorder in the affected *p*-tosylate ions has an kinetic onset and delay that is not directly proportional to that of MS1,[78] further complicating the modelling of nano-optomechanical transducers such as sample **IV**.

6.5 Conclusions

Structural changes from solid-state SO_2 linkage photo-isomerism in four ruthenium-sulfur dioxide based complexes have been investigated via pair distribution function analysis. Synchrotron-based X-ray diffraction provided the experimental data which were collected on finely powdered crystals, before and after sample photo-irradiation. Structural models of $G_{sim}(r)_{light}$ and $G_{sim}(r)_{dark}$ profiles were simulated from light-induced crystal structures that had been determined previously via photo-crystallography. $\Delta G(r)_{light-dark}$ profiles were used to isolate light-induced structural changes. These profiles showed good consistency between experimental data and the simulated structural models from photo-crystallography, especially in the 2-3 Å region. This is where the direct and second-neighbour pairwise atomic correlations associated with SO_2 linkage photo-isomerism are expected to be most evident. Statistical inference was employed to corroborate these findings. An average Pearson correlation coefficient of 0.94 was observed over this 2-3 Å region, showing that $\Delta G_{sim}(r)_{light-dark}$ profiles constructed from photo-crystallography results describe well the experimental data, $\Delta G_{exp}(r)_{light-dark}$, in this low-r region. Pearson correlation coefficients were lower in higher regions of r in $\Delta G(r)_{light-dark}$, especially for the sample (**IV**) which displays nano-optomechanical transduction. This poorer agreement indicated that some structural features were lacking in the photo-crystallography model of **IV**. Light-induced local atomic structure described exclusively by short-range order could be responsible for some of this deficit. This is because SO_2 photo-isomerisation may induce inhomogeneous spatial distributions of ruthenium-based cations that contain various mixes of dark state and photo-isomerised SO_2 configurations, while mechanical motion of some of the *p*-tosylate anions may mechanically move to avoid uncomfortably close proximity to the η^1-O-bound SO_2 photoisomer in **IV**. More generally, this study has shown that the single-crystal nature of the linkage photo-isomerism, already known in these samples, can persist once their crystals have been crushed into fine powders. Moreover, we have outlined an experimental procedure that paves the way for developing structural investigations on these types of materials using pair distribution function analysis.

7. Conclusions and Future Work

The work presented in thesis is centred around the application of radial and pair distribution function analysis to materials science problems, in the form of: an image-recognition based software tool that extracts information about particles from microscopy data; the development and application of a novel X-ray diffraction based experimental setup that affords PDF data on thin-film samples; and PDF and statistical correlation analysis, as applied to X-ray diffraction data that had been acquired on three types of complex materials.

The conclusions of these projects, along with suggested new avenues for further research that they have uncovered, are presented in the associated sections below.

7.1 ImageDataExtractor

ImageDataExtractor is a new software tool that auto-extracts microscopy images and analyses them quantitatively to extract information such as the shape, size and distribution of the particles contained in the image. It can be used as a stand-alone tool, on individual images, or as part of a larger framework which performs auto-extraction of relevant images from scientific documents.

In contrast to other image-recognition based tools designed for microscopy images, it is capable of autonomous, high-throughput operation. Calculation of evaluation metrics show that the tool performs well at all stages of image processing, and better than human error associated with manual image processing for the majority of its steps. The tool and cognate code are also open source.

The high-throughput image extraction and quantitative analysis options of *ImageDataExtractor* mean that this tool is inherently poised toward our ultimate goal of auto-generating material databases of information on particles.

Although *ImageDataExtractor* can provide useful, quantitative information regarding the particles in an image, its capabilities are currently limited to two dimensions. Expanding its capabilities so that it can draw conclusions in three dimensions could be done through the incorporation of depth cueing. Its accounting for 3-D factors would improve the accuracy of the calculated metrics of the tool. This may be accomplished by using the mean colour of the particles (a value already calculated in the *ImageDataExtractor* pipeline) to estimate the 3-D distance between given particles, since the pixel intensity is inversely related to 3-D depth.

Another possible improvement would be the training of an OCR library, specialised on fonts and characters used in imaging techniques, to use in scale recognition in order to improve the precision and recall of the scale detection step.

Lastly, after a sufficiently large particle database has been built, machine-learning techniques could be applied to help guide detection in new samples containing materials that are already in the database, and possibly even predict particle properties using similar compounds or synthesis processes from the materials database.

7.2 Glancing-angle PDF Analysis of DSSC Working Electrodes

The thin-film working electrodes of **N3-** and **N749-**sensitised DSSCs, comprising a dye···TiO_2 interface, were probed using PDF analysis of synchrotron-based X-ray diffraction data. The thin-film nature of this study required the development of a novel experimental setup for PDF data acquisition, glancing-angle PDF (gaPDF). ΔPDF analysis was employed to isolate the structure of this interface, by subtracting *D(r)* profiles collected on unsensitised DSSCs from those of sensitised DSSCs.

The resulting $\Delta D(r)^{exp}$ profiles, that represent the interfacial structure, were compared to $\Delta D(r)^{sim}$ profiles that were generated from simulated structural models of the interface using DFT calculations. Moderately good correlations with the simulations implied that, although we were able to observe the interface, our models require refinement. Closer examination of the deviations of the experimental data from those of the models revealed shorter-than-expected average bond lengths at values of r that are associated with anchoring groups. This suggests a prevalence of the BB mode, which features shorter anchoring bonds relative to the monodentate ester mode that is also included in the DFT model. This study demonstrates the successful application of the gaPDF method to DSSCs.

Moving forward, this method could be applied to all kinds of DSSCs, prepared with varying dyes or substrates. This would allow further insight to the structure of working electrodes in DSSCs, guiding subsequent research and development. Studies that use dyes which are known to anchor through different anchoring groups, such as pyridine, phosphonic acid or tetracyanate groups,[133] would be especially interesting for comparison of the interfacial structures that they form.

Additionally, this method is certainly not limited to the realm of DSSCs and could also be applied to a wide-range of thin-film samples.

7.3 Gd···Gd Separations in Amorphous $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$

ΔPDF analysis of AXS results on $(Gd_2O_3)_{0.230}(P_2O_5)_{0.770}$ were carried out in concert with various statistical techniques and corroboration with findings from other studies. This evidenced that rare-earth phosphate glasses show mixed ultra- and metaphosphate structural character, when their rare-earth composition lies between the two stoichiometric boundaries of RP_5O_{14} and $R(PO_3)_3$ [x = 0.167 and x = 0.250, respectively, according to the aforementioned formula $(R_2O_3)_x(P_2O_5)_{1-x}$]. The findings have confirmed previous peak assignments and related certain R···R correlations to archetypal meta- or ultraphosphate structural characteristics: 4.2(1)Å (meta),

4.8(1) Å (ultra/meta), 5.1(1) Å (ultra), 5.4(1) Å (meta), and 6.6(1) Å (meta). R···X correlations lying at average separations of 7.1(1), 7.6(1), 7.9(1), 8.4(1) and 8.7(1) Å were also resolved, thanks to the high spatial resolution enabled by anomalous X-ray scattering.

The majority of the well-resolved peaks in this study were observed in all four ΔPDFs that each represent the AXS signal, as determined by subtracting X-ray diffraction data collected at an X-ray energy that lies far from the Gd K-edge from a cognate data set acquired at an energy on the edge. Four X-ray energies were employed to obtain the diffraction data for these subtracted pairings, which lie at 10 eV, 40 eV, 300 eV and 500 eV from the Gd K-edge; ΔPDFs involving the K-10 eV data set afforded the most intense and reliable AXS signal, as expected.

These findings demonstrate exciting prospects for future AXS studies on other amorphous materials of similar nature.

7.4 Local Structure of SO_2 Photo-isomers in Ruthenium Coordination Complexes

Light-induced structural changes in four ruthenium-sulfur dioxide based complexes have been investigated via PDF analysis. Synchrotron-based X-ray diffraction data collected before and after sample photo-irradiation were compared to structural models of $G_{sim}(r)_{light}$ and $G_{sim}(r)_{dark}$ profiles that were simulated from previous photo-crystallographic results. $\Delta G(r)_{light-dark}$ profiles were used to isolate light-induced structural changes.

The profiles showed good consistency, especially in the 2-3 Å region where direct and second-neighbour pairwise atomic correlations associated with the SO_2 linkage photo-isomerism are expected to manifest. An average Pearson correlation coefficient of 0.94 across all four samples over the 2-3 Å region corroborated these results.

Poorer agreement in the higher-r regions, accompanied by lower Pearson correlation coefficients, indicated an absence of certain structural features in the photo-crystallographic models. Light-induced local atomic structure described exclusively by short-range order may be partially responsible for this disagreement. The disparities were found to be more severe for sample **IV,** which is known to display the phenomenon of nano-optomechanical transduction. The mechanical motion of some of the *p*-tosylate anions in this sample would contribute to such short-range order.

This study showed that the linkage photo-isomerism, already observed in these samples in single-crystal form, was also achievable in powder form. Additionally, this study outlines an experimental procedure paving the way for similar structural investigations on these types of compounds. Continuing investigations on a wider range of such materials, aided by more complex models that incorporate inter-ionic interactions over larger domains, are possible areas of further research in this direction.

Appendix

Appendix A1

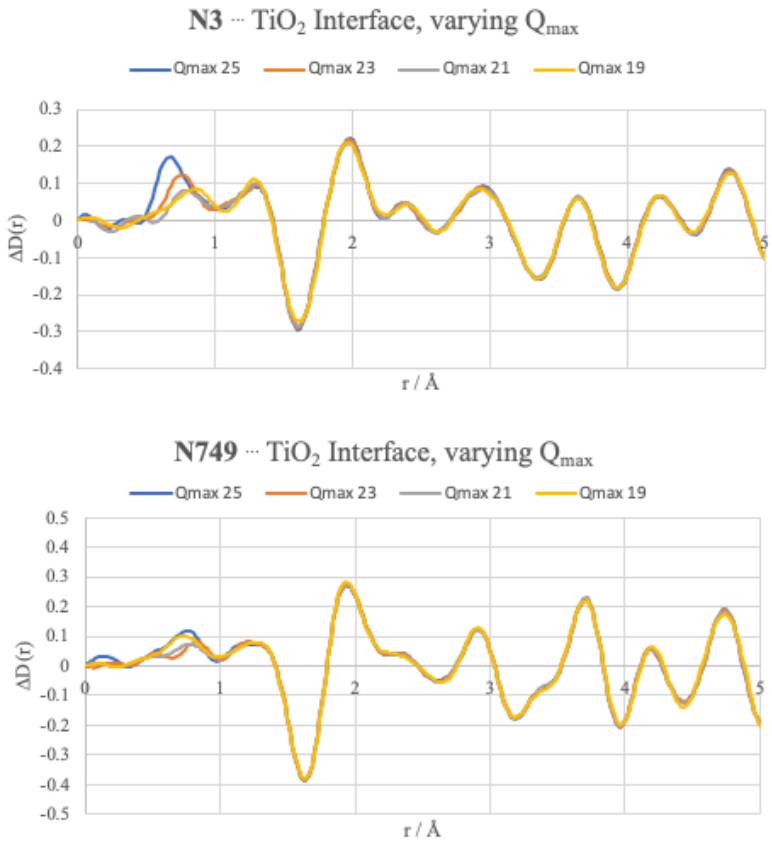


Figure A1.1 *Experimental ΔD(r) profiles, representing the dye···TiO_2 interface in* ***N3****- and* ***N749****-sensitised TiO_2 substrates with varying values of Q_{max}.*

Figure A1.1 shows the ΔD(r) representing the dye···TiO_2 interface in both cases, with varying Q_{max}. The limited effect that this testing imparts on the ΔD(r) shows how minimally the features were affected by Fourier transform artefacts, such as Fourier series termination errors.

Appendix A2

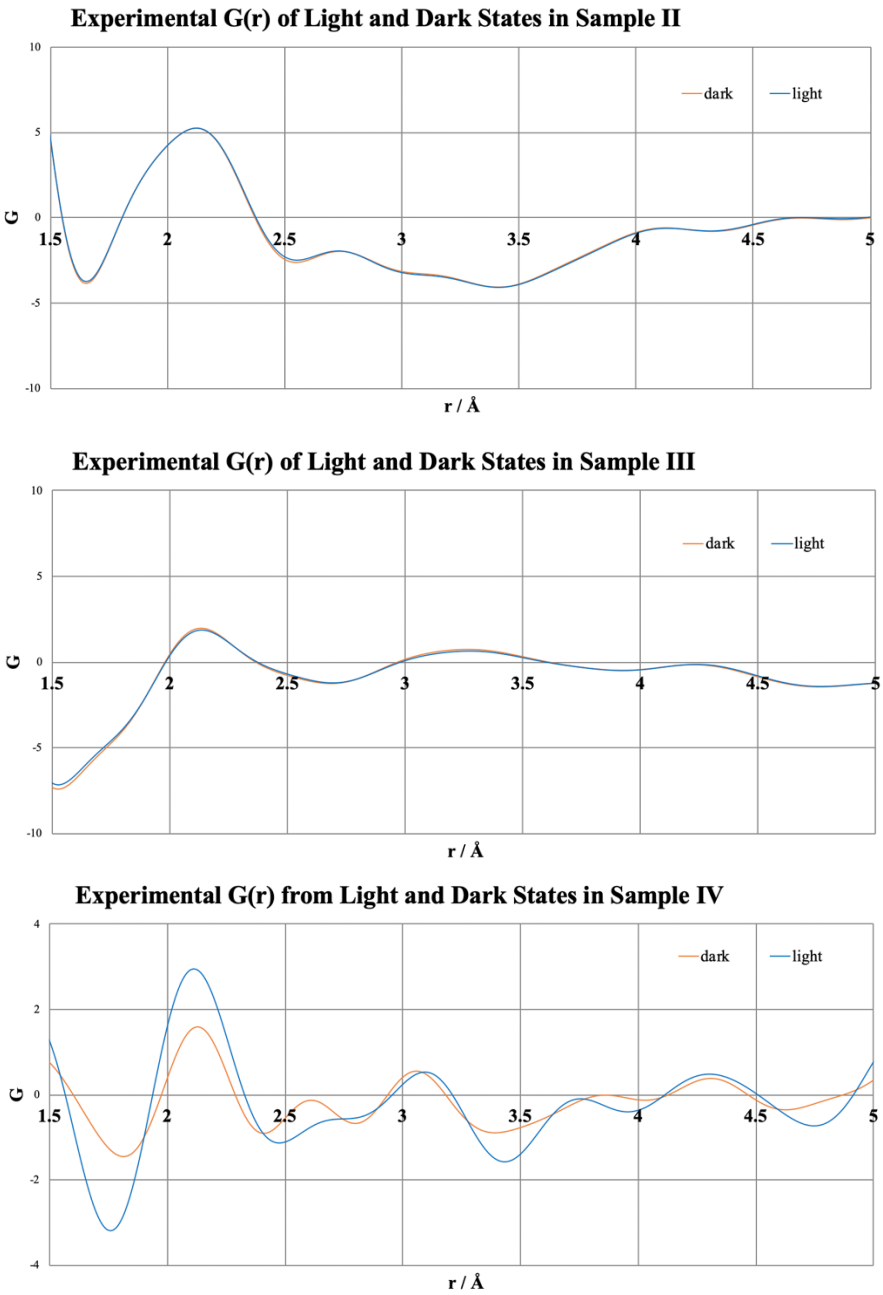


Figure A2.1 *The experimental G(r) profiles of the raw data of Samples* ***II-IV****, collected before and after sample photo-irradiation. The slightly different appearance of the G(r) for sample* ***IV*** *is attributed to the different software used for its processing.*

References

1. Billinge, S. J. The rise of the X-ray atomic pair distribution function method: a series of fortunate events. *Philosophical Transactions of the Royal Society A* **2019,** *377* (2147), 20180413.
2. Billinge, S. J. The nanostructure problem. *Physics* **2010,** *3,* 25.
3. Billinge, S. J.; Levin, I. The problem with determining atomic structure at the nanoscale. *Science* **2007,** *316* (5824), 561-565.
4. Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013,** *1* (1), 011002.
5. Hachmann, J.; Olivares-Amaya, R.; Jinich, A.; Appleton, A. L.; Blood-Forsythe, M. A.; Seress, L. R.; Roman-Salgado, C.; Trepte, K.; Atahan-Evrenk, S.; Er, S.; Shrestha, S. Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry–the Harvard Clean Energy Project. *Energy & Environmental Science* **2014,** *7* (2).
6. Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R. Efficient computational screening of organic polymer photovoltaics. *The Journal of Physical Chemistry Letters* **2013,** *4* (10), 1613-1623.
7. Cole, J. M.; Low, K. S.; Ozoe, H.; Stathi, P.; Kitamura, C.; Kurata, H.; Rudolf, P.; Kawase, T. Data mining with molecular design rules identifies new class of dyes for dye-sensitised solar cells. *Physical Chemistry Chemical Physics* **2014,** *16* (48), 26684-26690.
8. Cooper, C. B.; Beard, E. J.; Vázquez-Mayagoitia, Á.; Stan, L.; Stenning, G. B.; Nye, D. W.; Vigil, J. A.; Tomar, T.; Jia, J.; Bodedla, G. B.; Chen, S. Design-to-Device approach affords panchromatic co-sensitized solar cells. *Advanced Energy Materials* **2019,** *9* (5), 1802820.

9. Kim, E.; Huang, K.; Tomala, A.; Matthews, S.; Strubell, E.; Saunders, A.; McCallum, A.; Olivetti, E. Machine-learned and codified synthesis parameters of oxide materials. *Scientific Data* **2017,** *4,* 170217.

10. Kim, E.; Huang, K.; Saunders, A.; McCallum, A.; Ceder, G.; Olivetti, E. Materials synthesis insights from scientific literature via text extraction and machine learning. *Chemistry of Materials* **2017,** *29* (21), 9436-9444.

11. Kim, E.; Huang, K.; Jegelka, S.; Olivetti, E. Virtual screening of inorganic materials synthesis parameters with deep learning. *npj Computational Materials* **2017,** *3* (1), 53.

12. Tshitoyan, V.; Dagdelen, J.; Weston, L.; Dunn, A.; Rong, Z.; Kononova, O.; Persson, K. A.; Ceder, G.; Jain, A. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **2019,** *571* (7763), 95-98.

13. Swain, M. C.; Cole, J. M. ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. *Journal of Chemical Information and Modeling* **2016,** *56* (10), 1894-1904.

14. DeCost, B. L.; Lei, B.; Francis, T.; Holm, E. A. High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel. *Microscopy and Microanalysis* **2019,** *25* (1), 21-29.

15. Azimi, S. M.; Britz, D.; Engstler, M.; Fritz, M.; Mücklich, F. Advanced steel microstructural classification by deep learning methods. *Scientific Reports* **2018,** *8* (1), 2128.

16. Gola, J.; Webel, J.; Britz, D.; Guitar, A.; Staudt, T.; Winter, M.; Mücklich, F. Objective microstructure classification by support vector machine (SVM) using a combination of morphological parameters and textural features for low carbon steels. *Computational Materials Science* **2019,** *160,* 186-196.

17. Maksov, A.; Dyck, O.; Wang, K.; Xiao, K.; Geohegan, D. B.; Sumpter, B. G.; Vasudevan, R. K.; Jesse, S.; Kalinin, S. V.; Ziatdinov, M. Deep learning analysis of defect and phase evolution during electron beam-induced transformations in WS_2. *npj Computational Materials* **2019,** *5* (1), 12.

18. Roberts, G.; Haile, S. Y.; Sainju, R.; Edwards, D. J.; Hutchinson, B.; Zhu, Y. Deep learning for semantic segmentation of defects in advanced stem images of steels. *Scientific Reports* **2019,** *9* (1), 1-12.

19. Vlcek, L.; Maksov, A.; Pan, M.; Vasudevan, R. K.; Kalinin, S. V. Knowledge extraction from atomically resolved images. *ACS Nano* **2017,** *11* (10), 10313-10320.

20. O'Regan, B.; Grätzel, M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO2 films. *Nature* **1991,** *353* (6346), 737-740.

21. Bechinger, C.; Ferrere, S.; Zaban, A.; Sprague, J.; Gregg, B. A. Photoelectrochromic windows and displays. *Nature* **1996,** *383* (6601), 608.

22. Gong, J.; Sumathy, K.; Qiao, Q.; Zhou, Z. Review on dye-sensitized solar cells (DSSCs): Advanced techniques and research trends. *Renewable and Sustainable Energy Reviews* **2017,** *68,* 234-246.

23. Pagliaro, M.; Ciriminna, R.; Palmisano, G. BIPV: merging the photovoltaic with the construction industry. *Progress in Photovoltaics: Research and applications* **2010,** *18.1,* 61-72.

24. Zhang, K.; Qin, C.; Yang, X.; Islam, A.; Zhang, S.; Chen, H.; Han, L. High-performance, transparent, dye-sensitized solar cells for see-through photovoltaic windows. *Advanced Energy Materials* **2014,** *4* (11), 1301966.

25. Koops, S. E.; O'Regan, B. C.; Barnes, P. R.; Durrant, J. R. Parameters influencing the efficiency of electron injection in dye-sensitized solar cells. *Journal of the American Chemical Society* **2009,** *131* (13), 4808-4818.

26. Murakoshi, K.; Kano, G.; Wada, Y.; Yanagida, S.; Miyazaki, H.;

Matsumoto, M.; Murasawa, S. Importance of binding states between photosensitizing molecules and the TiO_2 surface for efficiency in a dye-sensitized solar cell. *Journal of Electroanalytical Chemistry* **1995,** *396* (1-2), 27-34.

27. De Angelis, F.; Fantacci, S.; Selloni, A.; Grätzel, M.; Nazeeruddin, M. K. Influence of the sensitizer adsorption mode on the open-circuit potential of dye-sensitized solar cells. *Nano letters* **2007,** *7* (10), 3189-3195.

28. Qin, P.; Yang, X.; Chen, R.; Sun, L.; Marinado, T.; Edvinsson, T.; Boschloo, G.; Hagfeldt, A. Influence of π-conjugation units in organic dyes for dye-sensitized solar cells. *The Journal of Physical Chemistry C* **2007,** *111* (4), 1853-1860.

29. Lee, K. E.; Gomez, M. A.; Elouatik, S.; Demopoulos, G. P. Further understanding of the adsorption mechanism of N719 sensitizer on anatase TiO_2 films for DSSC applications using vibrational spectroscopy and confocal Raman imaging. *Langmuir* **2010,** *26* (12), 9575-9583.

30. Jiang, X.; Karlsson, K. M.; Gabrielsson, E.; Johansson, E. M.; Quintana, M.; Karlsson, M.; Sun, L.; Boschloo, G.; Hagfeldt, A. Highly efficient solid-state dye-sensitized solar cells based on triphenylamine dyes. *Advanced Functional Materials* **2011,** *21* (15), 2944-2952.

31. Chen, H.; Gong, Y.; Vázquez-Mayagoitia, A.; Zhang, J.; Cole, J. M. Dye aggregation, photo-structural reorganization and multiple concurrent Dye…TiO_2 binding modes in dye-sensitized solar cell working electrodes containing benzothiadiazole-based dye RK-1. *ACS Applied Energy Materials* **2020,** *3* (1), 423-430.

32. Westermark, K.; Rensmo, H.; Siegbahn, H.; Keis, K.; Hagfeldt, A.; Ojamae, L.; Persson, P. PES Studies of $Ru(dcbpyH_2)_2(NCS)_2$ Adsorption on nanostructured ZnO for solar cell applications. *Journal of Physical Chemistry B*

2002, *106* (39), 10102-10107.

33. Rienzo, A.; Mayor, L. C.; Magnano, G.; Satterley, C. J.; Ataman, E.; Schnadt, J.; Schulte, K.; O'Shea, J. N. X-ray absorption and photoemission spectroscopy of zinc protoporphyrin adsorbed on rutile TiO_2 (110) prepared by in situ electrospray deposition. *The Journal of Chemical Physics* **2010,** *132* (8), 084703.

34. Pazoki, M.; Cappel, U. B.; Johansson, E. M.; Hagfeldt, A.; Boschloo, G. Characterization techniques for dye-sensitized solar cells. *Energy & Environmental Science* **2017,** *10* (3), 672-709.

35. Wagner, K.; Griffith, M. J.; James, M.; Mozer, A. J.; Wagner, P.; Triani, G.; Officer, D. L.; Wallace, G. G. Significant performance improvement of porphyrin-sensitized TiO_2 solar cells under white light illumination. *The Journal of Physical Chemistry C* **2010,** *115* (1), 317-326.

36. Griffith, M. J.; James, M.; Triani, G.; Wagner, P.; Wallace, G. G.; Officer, D. L. Determining the orientation and molecular packing of organic dyes on a TiO_2 surface using X-ray reflectometry. *Langmuir* **2011,** *27* (21), 12944-12950.

37. McCree-Grey, J.; Cole, J. M.; Evans, P. J. Preferred molecular orientation of coumarin 343 on TiO_2 surfaces: application to dye-sensitized solar cells. *ACS Applied Materials & Interfaces* **2015,** *7* (30), 16404-16409.

38. Cole, J. M.; Gong, Y.; McCree-Grey, J.; Evans, P. J.; Holt, S. A. Modulation of N3 and N719 dye$\cdots$ TiO_2 Interfacial structures in dye-sensitized solar cells as influenced by dye counter ions, dye deprotonation levels, and sensitizing solvent. *ACS Applied Energy Materials* **2018,** *1* (6), 2821-2831.

39. McCree-Grey, J.; Cole, J. M.; Holt, S. A.; Evans, P. J.; Gong, Y. Dye$\cdots$ TiO_2 interfacial structure of dye-sensitised solar cell working electrodes buried under a solution of I^-/I^{3-} redox electrolyte. *Nanoscale* **2017,** *9* (32), 11793-

11805.

40. Kley, C. S.; Dette, C.; Rinke, G.; Patrick, C. E.; Čechal, J.; Jung, S. J.; Baur, M.; Dürr, M.; Rauschenbach, S.; Giustino, F.; Stepanow, S.; Kern, K. Atomic-scale observation of multiconformational binding and energy level alignment of ruthenium-based photosensitizers on TiO_2 anatase. *Nano Letters* **2014,** *14* (2), 563-569.

41. Sasahara, A.; Pang, C.; Onishi, H. STM observation of a ruthenium dye adsorbed on a TiO_2 (110) surface. *The Journal of Physical Chemistry B* **2006,** *110* (10), 4751-4755.

42. Liu, K.; Hsu, C.; Ni, J.; Ho, K.; Lin, K. Photovoltaic properties of dye-sensitized solar cells associated with amphiphilic structure of ruthenium complex dyes. *Journak of Colloid and Interface Science.* **2012,** *372* (1), 73-79.

43. Marquet, P.; Andersson, G.; Snedden, A.; Kloo, L.; Atkin, R. Molecular scale characterization of the titania - dye - solvent interface in dye-sensitized solar cells. *Langmuir* **2010,** *26* (9), 9612-9616.

44. Debye, P.; H, M. The determination of the inner structure of liquids by X-ray means. *Phys Z* **1930,** *31,* 797-798.

45. Warren, B. E. X-ray determination of the structure of liquids and glass. *Journal of Applied Physics* **1937,** *8* (10), 645-654.

46. Kruh, R. F. Diffraction studies of the structure of liquids. *Chemical Reviews* **1962,** *62* (4), 319-346.

47. Furukawa, K. The radial distribution curves of liquids by diffraction methods. *Reports on Progress in Physics* **1962,** *25* (1), 395.

48. Warren, B. E. X-ray diffraction study of the structure of glass. *Chemical Reviews* **1940,** *26* (2), 237-255.

49. Fischer, H. E.; Barnes, A. C.; Salmon, P. S. Neutron and x-ray diffraction studies of liquids and glasses. *Reports on Progress in Physics* **2005,**

69 (1), 233.

50. Choi, J. J.; Yang, X.; Norman, Z. M.; Billinge, S. J.; Owen, J. S. Structure of methylammonium lead iodide within mesoporous titanium dioxide: active material in high-performance perovskite solar cells. *Nano Letters* **2013,** *14* (1), 127-133.

51. Jensen, K. M.; Blichfeld, A. B.; Bauers, S. R.; Wood, S. R.; Dooryhée, E.; Johnson, D. C.; Iversen, B. B.; Billinge, S. J. Demonstration of thin film pair distribution function analysis (tfPDF) for the study of local structure in amorphous and crystalline thin films. *IUcrJ* **2015,** *2* (5), 481-489.

52. Broer, M. M.; Bruce, A. J.; Grodkiewicz, W. H. Photoinduced refractive-index changes in several Eu $^{3+}$-, Pr $^{3+}$-, and Er $^{3+}$-doped oxide glasses. *Physical Review B* **1992,** *45* (13), 7077.

53. Carini, G.; D'angelo, G.; Tripodo, G.; Fontana, A.; Rossi, F.; Saunders, G. A. Low-energy magnetic excitations in the Pr metaphosphate glass. *Europhysics Letters* **1997,** *40* (4), 435.

54. Durville, F. M.; Behrens, E. G.; Powell, R. C. Laser-induced refractive-index gratings in Eu-doped glasses. *Physical Review B* **1986,** *34* (6), 4213.

55. Dong-Bing, H.; Chun-Lei, Y.; Ji-Meng, C.; Shun-Guang, L.; Li-Li, H. A novel Ce^{3+}/Tb^{3+} codoped phosphate glass as down-shifting materials for enhancing efficiency of solar cells. *Chinese Physics Letters* **2010,** *27* (11), 114208.

56. Liang, X.; Zhu, C.; Yang, Y.; Yuan, S.; Chen, G. Luminescent properties of Dy^{3+}-doped and Dy^{3+}–Tm^{3+} co-doped phosphate glasses. *Journal of Luminescence* **2008,** *128* (7), 1162-1164.

57. Saunders, G. A.; Brennan, T.; Acet, M.; Cankurtaran, M.; Senin, H. B.; Sidek, H. A. A.; Federico, M. Elastic and non-linear acoustic properties and thermal expansion of cerium metaphosphate glasses. *Journal of Non-crystalline*

Solids **2001,** *282* (2-3), 291-305.

58. Farok, H. M.; Senin, H. B.; Saunders, G. A.; Poon, W.; Vass, H. Optical and ultrasonic properties of europium phosphate glasses. *Journal of Materials Science* **1994,** *29* (11), 2847-2859.

59. Ford, P. J.; Graham, C. D.; Saunders, G. A.; Senin, H. B.; Cooper, J. R. Magnetic investigation of rare earth phosphate glasses. *Journal of Materials Science Letters* **1994,** *13* (10), 697-699.

60. Bowron, D. T.; Newport, R. J.; Rainford, B. D.; Saunders, G. A.; Senin, H. B. EXAFS and x-ray structural studies of $(Tb_2O_3)_{0.26}(P_2O_5)_{0.74}$ metaphosphate glass. *Physical Review B* **1995,** *51* (9), 5739.

61. Hoppe, U.; Kranold, R.; Stachel, D.; Barz, A.; Hannon, A. C. A neutron and X-ray diffraction study of the structure of the LaP_3O_9 glass. *Journal of Non-Crystalline Solids* **1998,** *232,* 44-50.

62. Cole, J. M.; van Eck, E. R.; Mountjoy, G.; Anderson, R.; Brennan, T.; Bushnell-Wye, G.; Newport, R. J.; Saunders, G. A. An x-ray diffraction and 31P MAS NMR study of rare-earth phosphate glasses,$(R_2O_3)_x(P_2O_5)_{1-x}$, x= 0.175-0.263, R= La, Ce, Pr, Nd, Sm, Eu, Gd, Tb, Dy, Ho, Er. *Journal of Physics: Condensed Matter* **2001,** *13* (18), 4105.

63. Hoppe, U.; Ebendorff-Heidepriem, H.; Neuefeind, J.; Bowron, D. T. A neutron and X-ray diffraction study of the structure of Nd phosphate glasses. *Zeitschrift für Naturforschung A* **2001,** *56* (3-4), 237-243.

64. Hoppe, U.; Metwalli, E.; Brow, R. K.; Neuefeind, J. High-energy X-ray diffraction study of La co-ordination in lanthanum phosphate glasses. *Journal of Non-Crystalline Solids* **2002,** *297* (2-3), 263-274.

65. Hoppe, U.; Brow, R. K.; Ilieva, D.; Jóvári, P.; Hannon, A. C. Structure of rare-earth phosphate glasses by X-ray and neutron diffraction. *Journal of Non-Crystalline Solids* **2005,** *351* (40-42), 3179-3190.

66. Hoppe, U.; Wyckoff, N. P.; Brow, R. K.; Rütt, U. Structure of ultraphosphate glasses with small rare-earth ions by X-ray diffraction. *Journal of Non-Crystalline Solids* **2011,** *357* (14), 2516-2521.

67. Cramer, A. J.; Cole, J. M.; FitzGerald, V.; Honkimaki, V.; Roberts, M. A.; Brennan, T.; Martin, R. A.; Saunders, G. A.; Newport, R. J. Effects of rare-earth co-doping on the local structure of rare-earth phosphate glasses using high and low energy X-ray diffraction. *Physical Chemistry Chemical Physics* **2013,** *15* (22), 8529-8543.

68. Cole, J. M.; van Eck, E. R.; Mountjoy, G.; Newport, R. J.; Brennan, T.; Saunders, G. A. A neutron diffraction and ^{27}Al MQMAS NMR study of rare-earth phosphate glasses,$(R_2O_3)_x (P_2O_5)_{1-x,}$ x= 0.187-0.263, R= Ce, Nd, Tb containing Al impurities. *Journal of Physics: Condensed Matter* **1999,** *11* (47), 9165.

69. Cole, J. M.; Newport, R. J.; Bowron, D. T.; Pettifer, R. F.; Mountjoy, G.; Brennan, T.; Saunders, G. A. A rare-earth K-edge EXAFS study of rare-earth phosphate glasses,$(R_2O_3)_x (P_2O_5)_{1-x}$, x= 0.187-0.239, R= La, Nd, Sm, Eu, Gd, Dy, Er. *Journal of Physics: Condensed Matter* **2001,** *13* (31), 6659.

70. Anderson, R.; Brennan, T.; Cole, J. M.; Mountjoy, G.; Pickup, D. M.; Newport, R. J.; Saunders, G. A. An extended x-ray absorption fine structure study of rare-earth phosphate glasses near the metaphosphate composition. *Journal of Materials Research* **1999,** *14* (12), 4706-4714.

71. Mountjoy, G.; Cole, J. M.; Brennan, T.; Newport, R. J.; Saunders, G. A.; Wallidge, G. W. A rare earth L3-edge EXAFS and L1-edge XANES study of Ce, Nd and Eu phosphate glasses and crystals in the composition range from metaphosphate to ultraphosphate. *Journal of Non-Crystalline Solids* **2001,** *279* (1), 20-27.

72. Bowron, D. T.; Saunders, G. A.; Newport, R. J.; Rainford, B. D.; Senin,

H. B. EXAFS studies of rare-earth metaphosphate glasses. *Physical Review B* **1996,** *53* (9), 5268.

73. Ilieva, D.; Kovacheva, D.; Cole, J. M.; Gutzow, I. Structure and devitrification chemistry of $Re(PO_3)_3$ (Re= La, Pr, Nd, Gd, Dy, Y) metaphosphate glasses. *Phosphorus Research Bulletin* **2002,** *13,* 137-146.

74. Cole, J. M. Applications of photocrystallography: a future perspective. *Zeitschrift für Kristallographie-Crystalline Materials* **2008,** *223* (4-5), 363-369.

75. Bochmann, J.; Vainsencher, A.; Awschalom, D. D.; Cleland, A. N. Nanomechanical coupling between microwave and optical photons. *Nature Physics* **2013,** *9* (11), 712.

76. Andréasson, J.; Straight, S. D.; Bandyopadhyay, S.; Mitchell, R. H.; Moore, T. A.; Moore, A. L.; Gust, D. Molecular 2: 1 digital multiplexer. *Angewandte Chemie International Edition* **2007,** *46* (6), 958-961.

77. Andréasson, J.; Straight, S. D.; Moore, T. A.; Moore, A. L.; Gust, D. Molecular all-photonic encoder− decoder. *Journal of the American Chemical Society* **2008,** *130* (33), 11122-11128.

78. Sylvester, S. O.; Cole, J. M. Solar-powered nanomechanical transduction from crystalline molecular rotors. *Advanced Materials* **2013,** *25* (24), 3324-3328.

79. Phillips, A. E.; Cole, J. M.; d'Almeida, T.; Low, K. S. Effects of the reaction cavity on metastable optical excitation in ruthenium-sulfur dioxide complexes. *Physical Review B* **2010,** *82* (15), 155118.

80. Kovalevsky, A. Y.; Bagley, K. A.; Cole, J. M.; Coppens, P. Light-induced metastable linkage isomers of ruthenium sulfur dioxide complexes. *Inorganic chemistry* **2003,** *42* (1), 140-147.

81. Sylvester, S. O.; Cole, J. M.; Waddell, P. G. Photoconversion bonding mechanism in ruthenium sulfur dioxide linkage photoisomers revealed by in situ

diffraction. *Journal of the American Chemical Society* **2012,** *134* (29), 11860-11863.

82. Sylvester, S. O.; Cole, J. M. Quantifying crystallographically independent optical switching dynamics in $RuSO_2$ photoisomers via lock-and-key crystalline environment. *The Journal of Physical Chemistry Letters* **2013,** *4* (19), 3221-3226.

83. Cole, J. M.; Velazquez-Garcia, J. D. J.; Gosztola, D. J.; Wang, S. G.; Chen, Y. S. Light-induced macroscopic peeling of single crystal driven by photoisomeric nano-optical switching. *Chemistry of Materials* **2019,** *31* (13), 4927-4935.

84. Bowes, K. F.; Cole, J. M.; Husheer, S. L.; Raithby, P. R.; Savarese, T. L.; Sparkes, H. A.; Teat, S. J.; Warren, J. E. Photocrystallographic structure determination of a new geometric isomer of $[Ru(NH_3)_4(H_2O)(\eta^1OSO)][MeC_6H_4SO_3]_2$. *Chemical Communications* **2006,** *23,* 2448-2450.

85. Phillips, A. E.; Cole, J. M.; d'Almeida, T.; Low, K. S. Ru–OSO coordination photogenerated at 100 K in tetraammineaqua (sulfur dioxide) ruthenium (II)(±)-camphorsulfonate. *Inorganic Chemistry* **2012,** *51* (3), 1204-1206.

86. Sylvester, S. O.; Cole, J. M.; Waddell, P. G.; Nowell, H.; Wilson, C. SO_2 phototriggered crystalline nanomechanical transduction of aromatic rotors in tosylates: Rationalization via photocrystallography of $[Ru(NH_3)_4SO_2X]$ $tosylate_2$ (X= pyridine, 3-Cl-pyridine, 4-Cl-pyridine). *The Journal of Physical Chemistry C* **2014,** *118* (29), 16003-16010.

87. Coppens, P.; Fomitchev, D. V.; Carducci, M. D.; Culp, K. Crystallography of molecular excited states. Transition-metal nitrosyl complexes and the study of transient species. *Journal of the Chemical Society, Dalton Transactions* **1998,** *6,* 865-872.

88. Cole, J. M. Single-crystal X-ray diffraction studies of photo-induced molecular species. *Chemical Society Reviews* **2004,** *33* (8), 501-513.

89. Cole, J. M. Photocrystallography. *Acta Crystallographica Section A: Foundations of Crystallography* **2008,** *64* (1), 259-271.

90. Cole, J. M. A new form of analytical chemistry: distinguishing the molecular structure of photo-induced states from ground-states. *Analyst* **2011,** *136* (3), 448-455.

91. Buschow, K. H. J. *Encyclopedia of Materials: Science and Technology;* Elsevier, 2001.

92. Soper, A. K.; Barney, E. R. Extracting the pair distribution function from white-beam X-ray total scattering data. *Journal of Applied Crystallography* **2011,** *44* (4), 714-726.

93. Keen, D. A. A comparison of various commonly used correlation functions for describing total scattering. *Journal of Applied Crystallography* **2001,** *34* (2), 172-177.

94. Egami, T.; Billinge, S. J. *Underneath the Bragg peaks: structural analysis of complex materials;* Elsevier, 2003.

95. Everitt, B.; Skrondal, A. *The Cambridge Dictionary of Statistics;* Cambridge University Press: Cambridge, 2002; Vol. 44.

96. Dykhne, T.; Taylor, R.; Florence, A.; Billinge, S. J. Data requirements for the reliable use of atomic pair distribution functions in amorphous pharmaceutical fingerprinting. *Pharmaceutical Research 28* (5), 1041-1048.

97. Davis, T.; Johnson, M.; Billinge, S. J. oward Phase Quantification at the Nanoscale Using the Total Scattering Pair Distribution Function (TSPDF) Method: Recrystallization of Cryomilled Sulfamerazine. *Crystal Growth & Design* **2013,** *13* (10), 4239-4244.

98. Terban, M. W.; Cheung, E. Y.; Krolikowski, P.; Billinge, S. J.

Recrystallization, phase composition, and local structure of amorphous lactose from the total scattering pair distribution function. *Crystal Growth & Design* **2015,** *16* (1), 210-220.

99. Tian, P.; Zhang, Y.; Senevirathne, K.; Brock, S. L.; Dixit, A.; Lawes, G.; Billinge, S. J. Diverse structural and magnetic properties of differently prepared MnAs nanoparticles. *ACS Nano* **2011,** *5* (4), 2970-2978.

100. Cole, J. M.; Cramer, A. J.; Shastri, S. D.; Mukaddem, K. T.; Newport, R. J. Multiple rare-earth ion environments in amorphous $(Gd_2O_3)_{0.230}$ $(P_2O_5)_{0.770}$ revealed by gadolinium K-edge anomalous x-ray scattering. *Physical Review Materials* **2018,** *2* (4), 045604.

101. Mukaddem, K. T.; Beard, E. J.; Yildirim, B.; Cole, J. M. ImageDataExtractor: A Tool to extract and quantify data from microscopy images. *Journal of Chemical Information and Modeling* **2019,** *Articles ASAP, doi: 10.1021/acs.jcim.9b00734*.

102. Holdren, J. P. Materials genome initiative for global competitiveness. Washingtonn, USA, 2011.

103. Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science* **2018,** *4* (2), 268-276.

104. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge structural database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2016,** *72* (2), 171-179.

105. Beard, E. J.; Sivaraman, G.; Vázquez-Mayagoitia, Á.; Vishwanath, V.; Cole, J. M. Comparative dataset of experimental and computational attributes of UV/vis absorption spectra. *Scientific Data* **2019,** *6* (1), 1-11.

106. Court, C. J.; Cole, J. M. Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction. *Scientific Data* **2018,** *5,* 180111.

107. Matas, J.; Galambos, C.; Kittler, J. Robust detection of lines using the progressive probabilistic hough transform. *Computer Vision and Image Understanding* **2000,** *78* (1), 119-137.

108. Zhang, Y.; Ye, Y.; Zhou, X.; Liu, Z.; Ma, D.; Li, B.; Liu, Q.; Zhu, G.; Chen, S.; Li, X. Facile preparation of a monodispersed CuO yolk-shelled structure with enhanced photochemical performance. *CrystEngComm* **2016,** *18* (41), 7994-8003.

109. Bradski, G.; Kaehler, A. *Learning OpenCV: Computer vision with the OpenCV library;* O'Reilly Media, Inc., 2008; Vol. 25, pp 120-125.

110. Suzuki, S. Topological structural analysis of digitized binary images by border following. *Computer Vision, Graphics, and Image Processing* **1985,** *30* (1), 32-46.

111. Xin, B.; Hao, J. Superhydrophobic self-assembled monolayers of long-chain fluorinated imidazolium ionic liquids. *RSC Advances* **2012,** *2* (12), 5141-5146.

112. PyPI Pytesseract Home Page. https://pypi.org/project/pytesseract (accessed Aug 22, 2019).

113. Smith, R. An overview of the Tesseract OCR engine. *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, 2007; pp 629-633.

114. Dong, C.; Loy, C. C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. *European conference on computer vision*, Cham, 2014; pp 184-199.

115. Dong, C.; Loy, C. C.; He, K.; Tang, X. Image super-resolution using

deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell* **2015,** *38,* 295-307.

116. LeCun, Y.; Boser, B.; Denker, J. S.; Henderson, D.; Howard, R. E.; Hubbard, W.; Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural computation* **1989,** *1* (4), 541-551.

117. Nair, V.; Hinton, G. E. Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010; pp 807-814.

118. Chollet, F. Xception: Deep learning with depthwise separable convolutions., 2017; pp 1251-1258.

119. Sezgin, M.; Sankur, B. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging* **2004,** *13* (1), 146-166.

120. Douglas, D. H.; Peucker, T. K. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization* **1973,** *10* (2), 112-122.

121. Shen, Z. R.; Wang, J. G.; Sun, P. C.; Ding, D. T.; Chen, T. H. Porous lanthanide oxides via a precursor method: Morphology control through competitive interaction of lanthanide cations with oxalate anions and amino acids. *Dalton Transactions* **2010,** *39* (26), 6112-6123.

122. Sklansky, J. Finding the convex hull of a simple polygon. *Pattern Recognition Letters* **1982,** *1* (2), 79-83.

123. Fitzgibbon, A. W.; Fisher, R. B. *A buyer's guide to conic fitting;* University of Edinburgh, Department of Artificial Intelligence, 1996. 513-522.

124. Hu, M. K. Visual pattern recognition by moment invariants. *IRE Transactions on Information Theory* **1962,** *8* (2), 179-187.

125. Helmlinger, J.; Sengstock, C.; Groß-Heitfeld, C.; Mayer, C.; Schildhauer, T. A.; Köller, M.; Epple, M. Silver nanoparticles with different size and shape: equal cytotoxicity, but different antibacterial effects. *RSC Advances* **2016,** *6* (22), 18490-18501.

126. Zhang, J.; Zhang, L.; Shen, X.; Xu, P.; Liu, J. Synthesis of BiOBr/WO_3 p–n heterojunctions with enhanced visible light photocatalytic activity. *CrystEngComm* **2016,** *18* (21), 3856-3865.

127. Yang, L. M.; Zhang, G. Y.; Wang, H. R.; Bai, X.; Shen, X. Q.; Liu, J. W.; Gao, D. Z. Mild synthesis of {001} facet predominated $Bi_2O_2CO_3$ clusters with outstanding simulated sunlight photocatalytic activities. *CrystEngComm* **2016**, *18*(20), 3683-3695.

128. Yoo, K.; Kim, J. Y.; Lee, J. A.; Kim, J. S.; Lee, D. K.; Kim, K.; Kim, J. Y.; Kim, B.; Kim, H.; Kim, W. M.; Kim, J. H. Completely transparent conducting oxide-free and flexible dye-sensitized solar cells fabricated on plastic substrates. *ACS Nano* **2015,** *9* (4), 3760-3771.

129. Wen, Z.; Yeh, M. H.; Guo, H.; Wang, J.; Zi, Y.; Xu, W.; Deng, J.; Zhu, L.; Wang, X.; Hu, C.; Zhu, L. Self-powered textile for wearable electronics by hybridizing fiber-shaped nanogenerators, solar cells, and supercapacitors. *Science Advances* **2016,** *2* (10), e1600097.

130. Yun, M. J.; Cha, S. I.; Seo, S. H.; Lee, D. Y. Insertion of dye-sensitized solar cells in textiles using a conventional weaving process. *Scientific Reports* **2015,** *5,* 11022.

131. Peng, M.; Yu, X.; Cai, X.; Yang, Q.; Hu, H.; Yan, K.; Wang, H.; Dong, B.; Zhu, F.; Zou, D. Waveguide fiber dye-sensitized solar cells. *Nano Energy* **2014,** *10,* 117-124.

132. Cole, J. M.; Pepe, G.; Al Bahri, O. K.; Cooper, C. B. Cosensitization in dye-sensitized solar cells. *Chemical Reviews* **2019,** *119* (12), 7279-7327.

133. Zhang, L.; Cole, J. M. Anchoring groups for dye-sensitized solar cells. *ACS Applied Materials & Interfaces* **2015,** *7* (6), 7279-7327.

134. Le Bahers, T.; Pauporté, T.; Laine, P. P.; Labat, F.; Adamo, C.; Ciofini, I. Modeling dye-sensitized solar cells: From theory to experiment. *The Journal of Physical Chemistry Letters* **2013,** *4* (6), 1044-1050.

135. Cole, J. M.; Low, K. S.; Gong, Y. Discovery of black dye crystal structure polymorphs: Implications for dye conformational variation in dye-sensitized solar cells. *ACS Applied Materials & Interfaces* **2015,** *7* (50), 27646-27653.

136. Low, K. S.; Cole, J. M.; Zhou, X.; Yufa, N. Rationalizing the molecular origins of Ru- and Fe-based dyes for dye-sensitized solar cells. *Acta Cryst* **2012,** *B68,* 137-149.

137. Cole, J. M.; Blood-Forsythe, M. A.; Lin, T. C.; Pattison, P.; Gong, Y.; Vázquez-Mayagoitia, A.; Waddell, P. G.; Zhang, L.; Koumura, N.; Mori, S. Discovery of S⋯ CN intramolecular bonding in a thiophenylcyanoacrylate-based dye: realizing charge transfer pathways and Dye⋯ TiO_2 anchoring characteristics for dye-sensitized solar cells. *ACS Applied Materials & Interfaces* **2014,** *9* (31), 25952-25961.

138. Zhang, F.; Ma, W.; Jiao, Y.; Wang, J.; Shan, X.; Li, H.; Lu, X.; Meng, S. Precise identification and manipulation of adsorption geometry of donor−π–acceptor dye on nanocrystalline TiO_2 films for improved photovoltaics. *ACS Applied Materials & Interfaces* **2014,** *6* (24), 22359-22369.

139. Deng, K.; Cole, J. M.; Rawle, J. L.; Nicklin, C.; Chen, H.; Yanguas-Gil, A.; Elam, J. W.; Stenning, G. B. G. Dye nanoaggregate structures in MK-2, N3, and N749 dye⋯TiO_2 interfaces that represent dye-sensitized solar cell working electrodes. *ACS Applied Energy Materials* **2019,** *Just Accepted, doi: 10.1021/acsaem.9b02002*.

140. Chen, H.; Cole, J. M.; Stenning, G. B. G.; Vazquez-Mayagoitia, A.; Yanguas-Gil, A.; Elam, J. W.; Stan, L.; Gong, Y. Imaging dye aggregation in MK-2, N3, N749 and SQ2 dye…TiO_2 interfaces that represent dye-sensitized solar cell working electrodes. *ACS Applied Energy Materials* **2019,** *Just Accepted doi:10.1021/acsaem.9b02002*.

141. Hagfeldt, A.; Boschloo, G.; Sun, L.; Kloo, L.; Pettersson, H. Dye-sensitized solar cells. *Chemical Reviews* **2010,** *110* (11), 6595-6663.

142. Pérez León, C.; Kador, L.; Peng, B.; Thelakkat, M. Characterization of the adsorption of Ru-bpy dyes on mesoporous TiO_2 films with UV− Vis, Raman, and FTIR spectroscopies. *The Journal of Physical Chemistry B* **2006,** *110* (17), 8723-8730.

143. Finnie, K. S.; Bartlett, J. R.; Woolfrey, J. L. Vibrational Spectroscopic study of the coordination of (2, 2 '-Bipyridyl-4, 4 '-dicarboxylic acid) ruthenium (II) complexes to the surface of nanocrystalline titania. *Langmuir* **1991,** *14* (10), 2744-2749.

144. Vittadini, A.; Selloni, A.; Rotzinger, F. P.; Grätzel, M. Formic acid adsorption on dry and hydrated TiO_2 anatase (101) surfaces by DFT calculations. *The Journal of Physical Chemistry B* **2000,** *104* (6), 1300-1306.

145. Soper, A. K. GudrunN and GudrunX: programs for correcting raw neutron and x-ray diffraction data to differential scattering cross section. *Technical Report RAL-TR-2011-013* **2011**.

146. Coelho, A. A. TOPAS and TOPAS-Academic: an optimization program integrating computer algebra and crystallographic objects written in C++. *Journal of Applied Crystallography* **2018,** *51* (1), 210-218.

147. Lorch, E. Neutron diffraction by germania, silica and radiation-damaged silica glasses. *Journal of Physics C: Solid State Physics* **1969,** *2* (2), 229.

148. Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.;

Van Dam, H. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; De Jong, W. A. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Computer Physics Communications* **2010,** *181* (9), 1477-1489.

149. Becke, A. D. Density-functional thermochemistry. III. the role of exact exchange. *The Journal of Chemical Physics* **1993,** *98* (7), 5648-5652.

150. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F. N.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force, fields. *The Journal of Physical Chemistry* **1994,** *98* (45), 11623-11627.

151. Hay, P. J.; Wadt, W. R. Ab initio effective core potentials for molecular calculations. Potentials for K to Au including the outermost core orbitals. **1985,** *82* (1), 299-310.

152. Jungsuttiwong, S.; Yakhanthip, T.; Surakhot, Y.; Khunchalee, J.; Sudyoadsuk, T.; Promarak, V.; Kungwan, N.; Namuangruk, S. The effect of conjugated spacer on novel carbazole derivatives for dye-sensitized solar cells: Density functional theory/time-dependent density functional theory study. *Journal of Computational Chemistry* **2012,** *33* (17), 1517-1523.

153. Jungsuttiwong, S.; Tarsang, R.; Sudyoadsuk, T.; Promarak, V.; Khongpracha, P.; Namuangruk, S. Theoretical study on novel double donor-based dyes used in high efficient dye-sensitized solar cells: the application of TDDFT study to the electron injection process. *Organic Electronics* **2013,** *14* (3), 711-722.

154. Tsai, H. H. G.; Tan, C. J.; Tseng, W. H. Electron transfer of squaraine-derived dyes adsorbed on TiO_2 clusters in dye-sensitized solar cells: A density functional theory investigation. *The Journal of Physical Chemistry C* **2015,** *119* (9), 4431-4443.

155. Kronig, R. D. L. On the theory of dispersion of x-rays. *Josa* **1926,** *12* (6), 547-557.

156. Kramers, H. A. La diffusion de la lumiere par les atomes. *Atti Cong. Intern. Fisica (Transactions of Volta Centenary Congress)*, 1927; pp 545-557.

157. Mierzejewski, A.; Saunders, G. A.; Sidek, H. A. A.; Bridge, B. Vibrational properties of samarium phosphate glasses. *Journal of Non-Crystalline Solids* **1988,** *104* (2-3), 323-332.

158. Liu, S. Z.; Hong, G. Y.; Hu, N. H. Crystal structure and ferroelastic domain of PrP_5O_{14}. *Acta Physica Sinica.* **1991,** *40* (1), 64-69.

159. Matuszewski, J.; Kropiwnicka, J.; Znamierowska, T. The crystal structure of lanthanum metaphosphate LaP_3O_9. *Journal of Solid State Chemistry* **1988,** *75* (2), 285-290.

160. Naili, H.; Ettis, H.; Yahyaoui, S.; Mhiri, T. Growth, structural, IR and magnetic studies of a new gadolinium metaphosphate, $Gd(PO_3)_3$. *Materials Research Bulletin* **2006,** *41* (11), 1990-1999.

161. Murashova, E. V.; Chudinova, N. N.; Ilyukhin, A. B. Specific features of the crystal structure of erbium polyphosphate of the structural type C. *Crystallography Reports* **2007,** *52* (2), 230-234.

162. Schulz, H.; Thiemann, K. H.; Fenner, J. The high-temperature phase of NdP_5O_{14}. *Materials Research Bulletin* **1974,** *9* (11), 1525-1530.

163. Cole, J. M.; Hannon, A. C.; Martin, R. A.; Newport, R. J. Direct observation of R… R distances in rare-earth (R) phosphate glasses by magnetic difference neutron diffraction. *Physical Review B* **2006,** *73* (10), 104210.

164. Wright, A. C.; Cole, J. M.; Newport, R. J.; Fisher, C. E.; Clarke, S. J.; Sinclair, R. N.; Fischer, H. E.; Cuello, G. J. The neutron diffraction anomalous dispersion technique and its application to vitreous $Sm_2O_3 \cdot 4P_2O_5$. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators,*

Spectrometers, Detectors and Associated Equipment **2007,** *571* (3), 622-635.

165. Naumov, P.; Chizhik, S.; Panda, M. K.; Nath, N. K.; Boldyreva, E. Mechanically responsive molecular crystals. *Chemical Reviews* **2015,** *115* (22), 12440-12490.

166. Abendroth, J. M.; Bushuyev, O. S.; Weiss, P. S.; Barrett, C. J. Controlling motion at the nanoscale: rise of the molecular machines. *ACS Nano* **2015,** *9* (8), 7746-7768.

167. Hammersley, A. P. FIT2D: an introduction and overview. *European Synchrotron Radiation Facility Internal Report ESRF97HA02T* **1997,** *68,* 58.

168. Yang, X.; Juhas, P.; Farrow, C. L.; Billinge, S. J. xPDFsuite: an end-to-end software solution for high throughput pair distribution function transformation, visualization and analysis. *arXiv preprint arXiv:1402.3163* **2014**.

169. Momma, K.; Izumi, F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of Applied Crystallography* **2011,** *44* (6), 1272-1276.

170. Fylstra, D.; Lasdon, L.; Watson, J.; Waren, A. Design and use of the Microsoft Excel Solver. *Interfaces* **1998,** *28* (5), 29-55.