ROYAL SOCIETY
OF CHEMISTRY

EDGE ARTICLE

Check for updates

A structure determination protocol based on combined analysis of 3D-ED data, powder XRD data, solid-state NMR data and DFT-D calculations reveals the structure of a new polymorph of L-tyrosine†

Christopher J. H. Smalley, [a] Harriet E. Hoskyns,[a] Colan E. Hughes, [a]
Duncan N. Johnstone,[b] Tom Willhammar, [c] Mark T. Young, [d]
Christopher J. Pickard,[be] Andrew J. Logsdail, [f] Paul A. Midgley[b]
and Kenneth D. M. Harris [*a]

We report the crystal structure of a new polymorph of L-tyrosine (denoted the β polymorph), prepared by crystallization from the gas phase following vacuum sublimation. Structure determination was carried out by combined analysis of three-dimensional electron diffraction (3D-ED) data and powder X-ray diffraction (XRD) data. Specifically, 3D-ED data were required for reliable unit cell determination and space group assignment, with structure solution carried out independently from both 3D-ED data and powder XRD data, using the direct-space strategy for structure solution implemented using a genetic algorithm. Structure refinement was carried out both from powder XRD data, using the Rietveld profile refinement technique, and from 3D-ED data. The final refined structure was validated both by periodic DFT-D calculations, which confirm that the structure corresponds to an energy minimum on the energy landscape, and by the fact that the values of isotropic ^{13}C NMR chemical shifts calculated for the crystal structure using DFT-D methodology are in good agreement with the experimental high-resolution solid-state ^{13}C NMR spectrum. Based on DFT-D calculations using the PBE0-MBD method, the β polymorph is meta-stable with respect to the previously reported crystal structure of L-tyrosine (now denoted the α polymorph). Crystal structure prediction calculations using the AIRSS approach suggest that there are three other plausible crystalline polymorphs of L-tyrosine, with higher energy than the α and β polymorphs.

Introduction

For many materials, it is difficult to prepare single-crystal specimens of suitable size to allow the measurement of single-crystal X-ray diffraction (XRD) data of adequate quality for crystal structure determination. Under such circumstances, crystal structure determination must rely on the use of other experimental approaches that allow diffraction data to be recorded on microcrystalline samples. One approach is to carry out structure determination directly from powder XRD data.[1–9] However, in spite of continuing advances in data measurement techniques and data analysis strategies, successful structure determination from powder XRD data may be rendered difficult or impossible due to certain intrinsic challenges that originate from the nature of the powder sample itself, for example: (i) if the powder sample contains two or more different crystalline phases (*e.g.*, containing an unknown impurity phase together with the main phase of interest), the process of unit cell determination may be very difficult or impossible, and (ii) if the powder sample exhibits a non-random distribution of crystallite orientations (so-called "preferred orientation"), the relative intensities of peaks in the experimental powder XRD data may differ significantly from the intrinsic relative diffraction intensities characteristic of the crystal structure – as a consequence, achieving successful structure solution from the powder XRD data may face insurmountable challenges.

[a]School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, Wales, UK. E-mail: HarrisKDM@cardiff.ac.uk

[b]Department of Materials Science, University of Cambridge, 27 Charles Babbage Road, Cambridge CB3 0FS, England, UK

[c]Department of Materials and Environmental Chemistry, Stockholm University, Svante Arrhenius väg 16C, 106 91 Stockholm, Sweden

[d]School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK

[e]Advanced Institute for Materials Research, Tohoku University 2-1-1 Katahira, Aoba, Sendai, 980-8577, Japan

[f]Cardiff Catalysis Institute, School of Chemistry, Cardiff University, Park Place, Cardiff CF10 3AT, Wales, UK

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1sc06467c

An alternative approach that circumvents both of these problems is to record three-dimensional electron diffraction (3D-ED) data on individual micro-crystals within the powder sample, effectively yielding single-crystal diffraction data for the selected micro-crystals. Recent progress in the development of instrumentation and data collection protocols has facilitated the measurement of 3D-ED data suitable for use in crystal structure determination calculations.[10–20] Nevertheless, the 3D-ED approach also suffers from certain intrinsic challenges, including the susceptibility of materials to undergo beam damage upon electron beam irradiation. This may result in data of low resolution and/or low completeness, which can hamper structure solution and refinement. Furthermore, intensities obtained using electron diffraction will be affected by dynamical scattering. Although structure solution and refinement based on the kinematical scattering approximation are able to deliver a structural model that is a good representation of the crystal structure,[21] obtaining a more accurate description of the crystal structure requires more rigorous consideration of the effects of dynamical scattering.[15,16,19]

Based on consideration of the various challenges discussed above, an attractive strategy for crystal structure determination of micro-crystalline materials is to record *both* 3D-ED data *and* powder XRD data, and then (i) to tackle unit cell determination using the 3D-ED data, (ii) to carry out structure solution using the 3D-ED data or the powder XRD data (or, ideally, using both datasets independently), and (iii) to achieve high-quality structure refinement using the powder XRD data (ensuring, *inter alia*, that all XRD-detectable phases in the micro-crystalline powder sample are properly taken into account). In the present paper, we demonstrate the successful application of this strategy to determine the structure of a new polymorph of L-tyrosine. Two other components of the strategy used in this work are: (i) to apply periodic DFT-D calculations to improve the quality of the structural model at various stages during the structure determination process, including to provide a rigorous validation[22–26] of the final structure obtained in Rietveld refinement from powder XRD data, and (ii) to assess the agreement between calculated solid-state NMR data, computed by DFT-D methodology for the final structure obtained in Rietveld refinement from the powder XRD data, and the corresponding experimental solid-state NMR data, again as part of the structure validation process.[26–34]

Among the 20 directly-encoded proteinogenic amino acids in their natural (enantiomerically pure) form, each member of this set has now had a crystal structure reported, following recent structure determinations of L-arginine,[35] L-tryptophan[36,37] and L-lysine[38] [we note that one member (glycine) of this set is non-chiral, while the other 19 amino acids are chiral and exist in nature only as the L-enantiomer]. Polymorphism is also known for some of these biologically important materials, with structurally characterized polymorphs reported for L-cysteine, L-glutamic acid, glycine, L-histidine, L-isoleucine, L-leucine, L-phenylalanine, L-proline, L-serine and L-tryptophan.

In the present paper, we report the preparation and structural characterization of a new polymorph of L-tyrosine (Fig. 1) obtained by crystallization from the gas phase, with structure
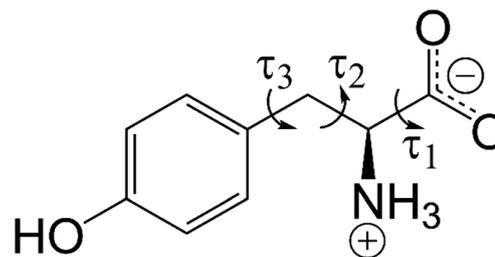

Fig. 1 Molecular structure of L-tyrosine in the zwitterionic form. The variable torsion angles (τ_1, τ_2 and τ_3) in the direct-space GA structure solution calculations are indicated.

determination carried out using the protocol outlined above, involving combined analysis of 3D-ED data and powder XRD data in conjunction with periodic DFT-D calculations, and with final structure validation also based on consideration of the agreement between calculated and experimental solid-state NMR data. Furthermore, the polymorphic landscape of L-tyrosine has been explored through crystal structure prediction calculations,[39–44] specifically using the AIRSS method.[45,46] The crystal structure of L-tyrosine determined previously (from single-crystal XRD data[47] and neutron diffraction data[48]) is now denoted the α polymorph, and the new polymorph reported here is denoted the β polymorph.

Results and discussion

Structure determination of the β polymorph of L-tyrosine

The new polymorph of L-tyrosine was prepared by crystallization from the gas phase following sublimation of a sample of the α polymorph (see *Methods*), and was clearly identified as a new solid form based on the fact that the powder XRD pattern is significantly different from that characteristic of the α polymorph. Evidence has been reported previously[49] for a new solid form of L-tyrosine prepared by this approach, but the crystal structure was not determined in the previous study (after completing the work in the present paper, it was apparent from the powder XRD data reported previously that the sample prepared by crystallization from the gas phase in the previous work[49] was a mixture of the α and β polymorphs).

Our initial attempts to progress with structure determination of the new polymorph from powder XRD data were unsuccessful as indexing (unit cell determination) using algorithms in the CRYSFIRE suite[50] could not account for all peaks present in the powder XRD pattern. However, 3D-ED data recorded for the same sample were indexed successfully using REDp[51] and subsequently XDS[52] to give the following unit cell with monoclinic metric symmetry: $a = 7.92$ Å, $b = 6.13$ Å, $c = 9.90$ Å, $\beta = 94.82°$, $V = 478.9$ Å^3. The space group was assigned from the 3D-ED data as $P2_1$, based on systematic extinctions. Given the volume of this unit cell, density considerations suggest that there are two molecules in the unit cell and therefore one molecule in the asymmetric unit for space group $P2_1$.

The unit cell determined from the 3D-ED data shed light on the failure to index the powder XRD data. Most peaks in the

powder XRD data were consistent with the unit cell determined from the 3D-ED data, but the powder XRD data also contained a few additional peaks due to a second crystalline phase (identified as the α polymorph of L-tyrosine; Fig. S2†). Clearly, the biphasic nature of the powder sample was the reason for the failure to index the powder XRD data in this case (however, we note that the presence of a minor phase in a biphasic powder sample does not always prevent successful unit cell determination of the major phase in attempting to index the powder XRD data; in favourable cases, it can be possible to index the peaks due to the major phase in a biphasic powder sample, as demonstrated in previous studies[37]).

Structure solution of the new polymorph of L-tyrosine was carried out using the 3D-ED data, as described below. Furthermore, although the powder XRD data contained additional peaks due to a second phase, the number and intensities of these peaks were relatively low, such that it was considered viable also to attempt structure solution from the powder XRD data (encouraged by previous cases in which structure solution from powder XRD data containing a small number of peaks due to an impurity phase were successful[37]). Thus, structure solution was also attempted from the powder XRD data.

Structure solution was carried out independently from the 3D-ED data and powder XRD data using the direct-space strategy implemented using a genetic algorithm (GA) in the program EAGER.[53–56] While EAGER was originally developed and applied for structure solution from powder XRD data,[57–63] it has been extended recently to allow structure solution from 3D-ED data within the kinematical scattering approximation.[64] In the direct-space GA structure solution calculations, the contents of the asymmetric unit comprised one molecule of L-tyrosine in the zwitterionic form, constructed using standard bond lengths and bond angles (based on geometric information from MOGUL[65] and, for bonds involving hydrogen, from Allen et al.[66]). Trial crystal structures were defined by a total of eight structural variables (two positional, three orientational and three torsional variables; the torsion-angle variables are defined in Fig. 1). We note that, for space group $P2_1$, the position of the molecule along the b-axis can be fixed; thus, only two positional variables are required in the direct-space search.

In the structure solution calculations using the 3D-ED data and the powder XRD data (for the original biphasic sample), a total of 40 independent GA structure solution calculations were carried out in each case from different random initial populations. Each GA calculation involved the evolution of a population of 100 trial structures for 100 generations, with 10 mating operations and 50 mutation operations carried out per generation. The R-factors used to assess the quality of trial structures (based on the level of agreement between calculated and experimental diffraction data) in the structure solution calculations from the 3D-ED data (R_F) and powder XRD data (R_{wp}) are defined in Section S4,† and the evolution of lowest R-factor in the population as a function of generation number in the 40 independent GA calculations using the 3D-ED data and the powder XRD data are shown in Fig. S4.† Using the 3D-ED data, 7 of the 40 independent GA calculations generated essentially the same structure solution with lowest R_F, whereas
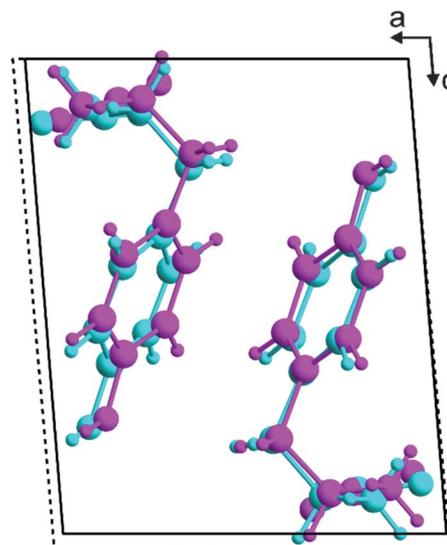

Fig. 2 Overlay of the best structure solutions (with lowest R-factor) viewed along the b-axis for the β polymorph of L-tyrosine obtained from the direct-space GA structure solution calculations using 3D-ED data (cyan; unit cell shown by the dashed black lines) and powder XRD data (magenta; unit cell shown by the solid black lines).

using the powder XRD data, 38 of the 40 independent GA calculations generated essentially the same structure solution with lowest R_{wp} (Fig. S4†). The best structure solutions (with lowest R-factor) from the 3D-ED data and the powder XRD data represent essentially the same structure (Fig. 2), with root-mean-squared difference (RMSD) of only 0.33 Å between the positions of the non-hydrogen atoms in these structures. The best structure solution from the powder XRD data was used as the starting structural model for Rietveld refinement,[67] which was carried out using TOPAS.[68]

Our initial Rietveld refinement used the powder XRD dataset for the original biphasic sample that was used in the structure solution calculations described above. However, at this stage, further attempts to prepare the new polymorph of L-tyrosine by crystallization from the gas phase under different experimental conditions (see Methods) were successful in producing a monophasic sample, confirmed by powder XRD. A high-quality powder XRD dataset for this monophasic sample was used for the final Rietveld refinement. Initially, profile fitting using the Pawley method[69] gave a good fit to the experimental powder XRD data (Fig. 3; R_{wp} = 0.29%, R_p = 0.22%). The structure solution from the EAGER calculations on the powder XRD data discussed above (Fig. 2) was used as the initial structural model for the Rietveld refinement, with standard restraints applied to bond lengths and bond angles, and planar restraints applied to the phenyl ring and carboxylate group. A common isotropic displacement parameter was refined for all non-hydrogen atoms, and the isotropic displacement parameter for hydrogen atoms was set at 1.2 times this value. As powder XRD data recorded for the same powder sample using a two-dimensional detector (Fig. S3†) clearly showed non-uniform intensity of the Debye–Scherrer rings, it was concluded that
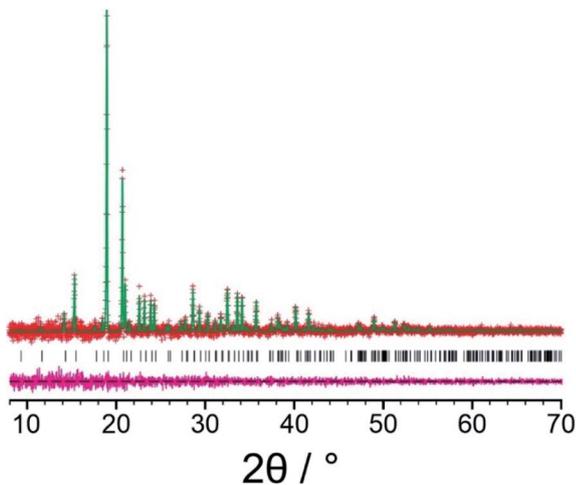
Fig. 3 Profile fitting (using the Pawley method) of the powder XRD data (background subtracted) for the β polymorph of L-tyrosine, showing the experimental powder XRD data (red "+" marks), the calculated powder XRD data (green line), the predicted peak positions (black tick marks) and the difference between experimental and calculated powder XRD data (magenta line).
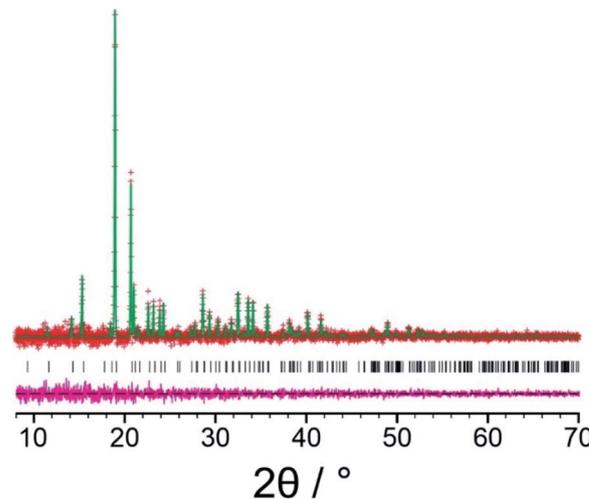

Fig. 4 Final Rietveld refinement of the powder XRD data (background subtracted) for the β polymorph of L-tyrosine, showing the experimental powder XRD data (red "+" marks), the calculated powder XRD data (green line), the predicted peak positions (black tick marks) and the difference between experimental and calculated powder XRD data (magenta line).

the powder sample exhibits preferred orientation; thus, a preferred orientation correction was applied in the Rietveld refinement using the March–Dollase method,[70,71] with a refined March parameter of 0.67 for the (0 1 0) plane. After completing the Rietveld refinement, the refined structure was subjected to DFT-D geometry optimization with fixed unit cell, which led to only minor structural changes (we emphasize that, in utilizing DFT-D geometry optimization in this way to facilitate improvements in the structural model in tandem with Rietveld refinement, it is crucial that the unit cell is fixed in the DFT-D geometry optimization calculations in order that the unit cell in the geometry-optimized structure matches the unit cell corresponding to the experimental powder XRD data). The geometry optimized structure was then used as the starting model for the final Rietveld refinement, which included additional restraints on the geometries of hydrogen bonds based on the structure obtained in the DFT-D geometry optimization. The final Rietveld refinement gave a good-quality fit to the powder XRD data (Fig. 4; $R_{wp} = 0.30\%$, $R_p = 0.23\%$), comparable to the quality of fit obtained in Pawley fitting (Fig. 3), with the following final refined parameters: $a = 7.6764(6)$ Å, $b = 5.8870(3)$ Å, $c = 9.6143(6)$ Å, $\beta = 94.575(4)°$, $V = 433.10(5)$ Å^3 (2θ range, 8°–70°; 3858 profile points; 93 refined variables; 81 restraints). The crystal structure of the β polymorph of L-tyrosine obtained in the final Rietveld refinement has been deposited in the CSD (deposition number: 2114085).

Least-squares refinement was also carried out against the 3D-ED data using the kinematical treatment. The starting structural model for the refinement was the structure solution obtained from the 3D-ED data using EAGER (see Fig. 2), with unit cell parameters taken from the profile-fitting (in this case Pawley fitting) of the powder XRD data and with hydrogen atoms included based on geometric restraints. An initial

refinement using isotropic atomic displacement parameters converged with $R1 = 25.1\%$. A final refinement using additional restraints on the C–C, C–O and C–N distances in order to improve the bonding geometries converged with $R1 = 25.1\%$. The final refined structures from the powder XRD data and 3D-ED data are in good agreement (Fig. S5;† RMSD for non-hydrogen atoms, 0.135 Å).

As further validation of the crystal structure of the β polymorph of L-tyrosine, DFT-D geometry optimization (using PBE-TS and with fixed unit cell) on the final refined structure from the Rietveld refinement led to only minor atomic displacements (Fig. S6;† RMSD for non-hydrogen atoms, 0.025 Å), confirming that the final refined crystal structure is very close to a minimum on the energy landscape. Furthermore, the isotropic solid-state ^{13}C NMR chemical shifts calculated using DFT-D/GIPAW methodology (see *Methods*) for the crystal structure of the β polymorph (Table S5†) are in close agreement with the isotropic chemical shifts in the experimental high-resolution solid-state ^{13}C NMR spectrum of this material, as shown in Fig. 5. Thus, in addition to being in excellent agreement with the experimental powder XRD data, as presented above (Fig. 4), our reported crystal structure of the β polymorph of L-tyrosine is also confirmed to be in excellent agreement with the experimental solid-state ^{13}C NMR data.

We note that the high-resolution solid-state ^{13}C NMR spectrum of the β polymorph (Fig. 5) is clearly distinct from that of the α polymorph (Fig. S10†), and in each case there is good agreement between the experimental solid-state ^{13}C NMR spectrum and the isotropic ^{13}C NMR chemical shifts calculated from the crystal structure using DFT-D/GIPAW methodology (Table S5†). These observations highlight the utility of DFT-D/GIPAW calculations to compute reliable solid-state NMR data
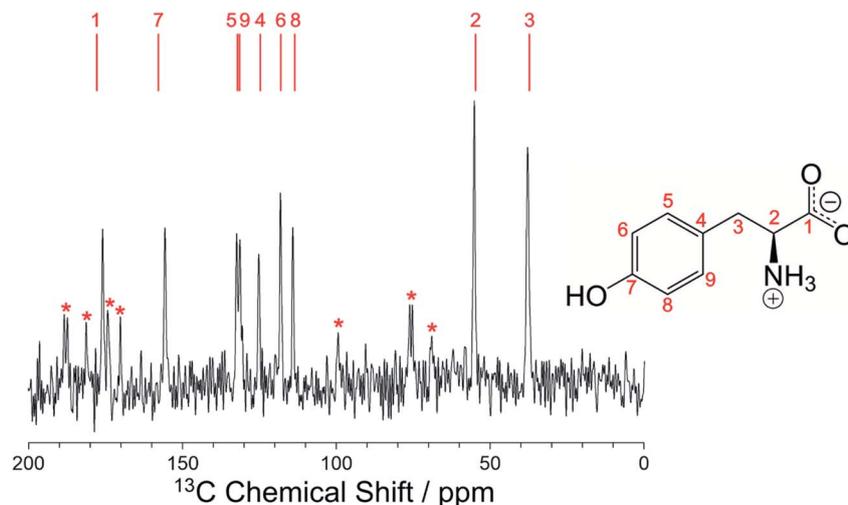
Fig. 5 Experimental high-resolution solid-state ^{13}C NMR spectrum recorded for the β polymorph of L-tyrosine together with the values of isotropic ^{13}C NMR chemical shifts calculated for the crystal structure of the β polymorph (indicated by the red lines above the spectrum). The specific ^{13}C site corresponding to each calculated value is indicated. Spinning sidebands in the experimental spectrum are marked by red asterisks.

for known crystal structures, both in the context of polymorph characterization and as a powerful strategy in conjunction with structure determination from powder XRD data.

Relative energies of the α and β polymorphs of L-tyrosine

To assess the relative energetic properties of the crystal structure of the β polymorph of L-tyrosine reported here and the crystal structure of the α polymorph reported previously[47] (from single-crystal XRD), periodic DFT-D calculations (see *Methods*) were carried out using different exchange-correlation functionals (PBE and PBE0) combined with different methods for dispersion correction (TS and MBD). For each polymorph, the experimentally determined crystal structure at ambient temperature was taken as the initial structural model for DFT-D geometry optimization (with fixed unit cell) using PBE-TS. For the resulting geometry optimized structures, single-point energy calculations were carried out using each of the four combinations of functional and dispersion correction method discussed above. For each calculation method, the energy is higher for the β polymorph than the α polymorph, with the following energy differences (expressed per mole of L-tyrosine molecules): 4.74 kJ mol^{-1} (PBE-TS), 3.96 kJ mol^{-1} (PBE-MBD), 3.66 kJ mol^{-1} (PBE0-TS), 3.64 kJ mol^{-1} (PBE0-MBD). Among these methods, PBE0-MBD is considered[72,73] to give the most reliable assessment of the relative energies of polymorphs of organic materials. We note that the calculated energy difference using PBE0-MBD of 3.64 kJ mol^{-1} between the α and β polymorphs is fully consistent[74] with the β polymorph being an experimentally observable meta-stable polymorph of L-tyrosine. From the crystal structures determined at ambient temperature, the densities of the α and β polymorphs are 1.414 g cm^{-3} and 1.389 g cm^{-3}, respectively. Clearly, the higher density of the α polymorph may be a contributing factor to the lower energy of this polymorph established from the DFT-D calculations.
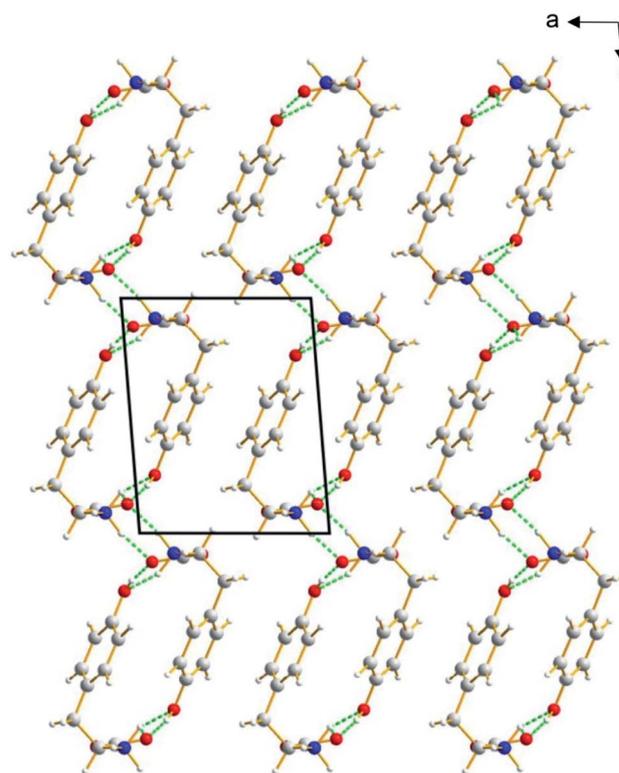

Fig. 6 Crystal structure of the β polymorph of L-tyrosine viewed along the *b*-axis, showing the alternating hydrophilic and hydrophobic layers parallel to the *ab*-plane (horizontal). The hydrogen-bonded ribbons involving the amino-acid head-groups propagate along the direction of view. Hydrogen bonds are indicated by green dashed lines.

Structural properties of the β polymorph of L-tyrosine

The β polymorph of L-tyrosine (Fig. 6) comprises alternate hydrophobic and hydrophilic layers parallel to the *ab*-plane.
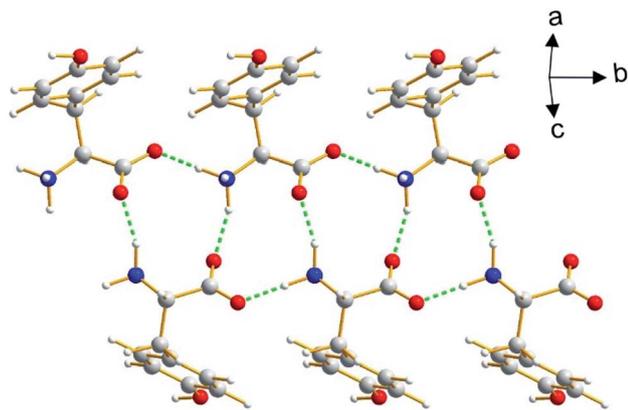
Fig. 7 The hydrogen-bonded ribbon involving the amino acid head-groups in the β polymorph of L-tyrosine. Hydrogen bonds are indicated by green dashed lines.

The hydrophilic region contains the amino acid head-groups and the OH groups of the side-chains, while the hydrophobic region contains the phenyl rings of the side-chains. The hydrogen-bonding involving the amino acid head-groups gives rise to a ribbon motif (Fig. 7) propagating along the b-axis. The ribbon is constructed from two strands of L-tyrosine molecules, with N–H···O hydrogen bonding between adjacent molecules in a given strand and between molecules in the two strands. The ribbon is also engaged in hydrogen bonding with the OH groups of the side-chains of the molecules that form the hydrogen-bonded ribbons in the layers "above" and "below" along the c-axis (Fig. 8). The contiguous hydrogen-bonded network comprises corrugated slabs (with a mean plane parallel to the bc-plane; Fig. 6) constructed from the hydrogen-bonded ribbons parallel to the b-axis and hydrogen-bonding (involving the OH groups) to adjacent ribbons along the c-axis. Adjacent corrugated slabs are related by translation along the a-axis and interact through van der Waals interactions.

We now compare the structural properties of the α and β polymorphs, firstly noting that the molecular conformation is similar in each polymorph. However, while the β polymorph contains one-dimensional hydrogen-bonded ribbons, the hydrophilic region of the α polymorph comprises a two-
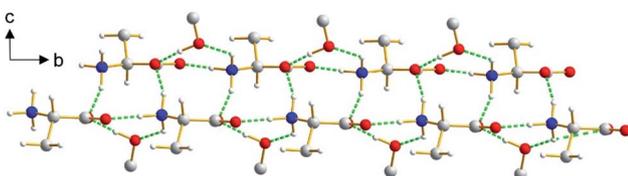

Fig. 8 Hydrogen bonding between the ribbon containing the amino acid head-groups and the OH groups of the side-chains (from molecules above and below the ribbon shown) in the β polymorph of L-tyrosine. For clarity, only the $CH_2CH(NH_3^+)CO_2^-$ unit of each head-group and only the COH unit of each side-chain are shown. The two strands of the ribbon propagate along the b-axis (horizontal) and are offset slightly from each other along the c-axis (vertical). Hydrogen bonds are indicated by green dashed lines.

dimensional hydrogen-bonded array (Fig. S8 and S9†) involving the amino acid head-groups and the OH groups of molecules in the layers above and below. Consequently, the α polymorph is a three-dimensionally connected hydrogen-bonded structure, in contrast to the two-dimensionally connected hydrogen-bonding arrangement (corrugated slabs) in the β polymorph.

A more detailed description of the crystal structure of the β polymorph, and comparison to the structural properties of the α polymorph, is given in Section S2.†

Crystal structure prediction of L-tyrosine

To further explore the polymorphic landscape of L-tyrosine, structure prediction calculations were carried out using the AIRSS methodology[45,46] (see *Methods*). The "initial" geometry optimization in AIRSS produced seven trial crystal structures for further consideration. One pair of trial structures (ranked 1 and 3 in the initial energy ranking using PBE-TS; see Table S1†) resembled the α polymorph and another pair of trial structures (ranked 2 and 6 in the initial energy ranking using PBE-TS; Table S1†) resembled the β polymorph.

The seven trial structures from AIRSS were then subjected to "precise" geometry optimization (using PBE-TS and including relaxation of the unit cell) in FHI-aims, and the results are summarized in Table S2.† In the "precise" geometry optimization, trial structures 1 and 3 converged (with an energy difference of only 0.04 kJ mol⁻¹) on the same structure corresponding to the α polymorph, while trial structures 2 and 6 converged (with an energy difference of only 0.06 kJ mol⁻¹) on the same structure corresponding to the β polymorph. The good agreement between the crystal structure corresponding to the β polymorph from this calculation (specifically structure 6, which had slightly lower energy than structure 2; see Table S2†) and the crystal structure of the β polymorph determined from Rietveld refinement of powder XRD data is shown in Fig. S7.† Trial structures 4, 5 and 7 represent three distinct polymorphs that have not been reported previously in experimental studies and are hereafter denoted "predicted structures" A, B and C, respectively.

Thus, the seven initial trial structures from AIRSS were reduced to five distinct structures, representing five predicted polymorphs of L-tyrosine. Following the "precise" geometry optimization, the energy of each of the five polymorphs was then determined using a single-point PBE0-MBD calculation. The structure corresponding to the α polymorph is lowest in energy and the structure corresponding to the β polymorph is next lowest in energy. The calculated PBE0-MBD energies of each polymorph, relative to the energy of the α polymorph, are: 4.10 kJ mol⁻¹ (β polymorph); 11.11 kJ mol⁻¹ (predicted structure A); 11.79 kJ mol⁻¹ (predicted structure B); 27.56 kJ mol⁻¹ (predicted structure C). As predicted structures A, B and C are all significantly higher in energy than the α polymorph, it is perhaps unlikely[74] that they represent experimentally accessible meta-stable polymorphs of L-tyrosine. However, the computational approach used here neglects entropic factors, which could have a significant influence on the relative energetic

properties of the different structures. Crystallographic data for the five structures predicted by AIRSS (following "precise" geometry optimization) are summarized in Table S3,† and cif files of these structures are included in ESI.†

It is interesting to note that predicted structure A (Fig. S11†) and predicted structure C (Fig. S13†) share several features in common with the β polymorph of L-tyrosine. In particular, these structures have a similar hydrogen-bonded ribbon motif (see Fig. 7) comprising two strands constructed from amino acid head-groups, with N–H⋯O hydrogen bonding between adjacent molecules in a given strand and between molecules in the two strands, and the hydrogen-bonded ribbon is also engaged in hydrogen bonding with the OH groups of molecules involved in adjacent hydrogen-bonded ribbons. The contiguous hydrogen-bonded network in these structures comprises corrugated slabs of L-tyrosine molecules [with mean plane parallel to the bc-plane for the β polymorph (Fig. 6) and parallel to the ab-plane for predicted structures A (Fig. S11†) and C (Fig. S13†)], which interact with adjacent corrugated slabs through van der Waals interactions. The structures of the β polymorph and predicted structures A and C differ in the details of the topology and the relative arrangement of these corrugated slabs.

In contrast, predicted structure B (Fig. S12†) contains hydrogen-bonded chains of L-tyrosine molecules linked by N–H⋯O interactions between amino acid head-groups (analogous to a single strand of the hydrogen-bonded ribbon motif observed in the β polymorph). Each hydrogen-bonded chain is also engaged in hydrogen bonding with the OH groups of molecules in an adjacent chain. In this structure, the contiguous hydrogen-bonded network comprises columns of molecules (extending along the a-axis; Fig. S12†), which interact with adjacent columns through van der Waals interactions.

Concluding remarks

In conclusion, a new polymorph of L-tyrosine (β polymorph) has been prepared by crystallization from the gas phase, with the crystal structure determined successfully by combined analysis of 3D-ED data and powder XRD data, in conjunction with periodic DFT-D calculations. The final structure of the β polymorph from Rietveld refinement was further validated by: (i) periodic DFT-D geometry optimization, which confirms that the crystal structure is very close to a minimum on the energy landscape, and (ii) comparison between isotropic ^{13}C NMR chemical shifts calculated for the crystal structure using DFT-D methodology and those observed in the experimental high-resolution solid-state ^{13}C NMR spectrum, which shows an excellent match between calculated and experimental values. Furthermore, DFT-D calculations at the PBE0-MBD level indicate that the crystal structure determined here for the β polymorph is meta-stable relative to the crystal structure of the α polymorph reported previously.

Structure prediction calculations using the AIRSS methodology generated five plausible polymorphic crystal structures for L-tyrosine. The lowest energy structure is the α polymorph, with the β polymorph higher in energy by ca. 4 kJ mol^{-1}; the other three predicted structures are higher in energy than the α polymorph by more than 11 kJ mol^{-1}.

In addition to expanding our understanding of the structural properties of L-tyrosine in the crystalline state, the research strategy described in this paper serves to highlight the significant advantages that can be gained by combined analysis of 3D-ED data and powder XRD data, together with the application of periodic DFT-D calculations and consideration of solid-state NMR data, within a robust protocol for crystal structure determination of materials that are unsuitable for structural characterization by single-crystal XRD.

Methods

Sample preparation

The β polymorph of L-tyrosine was prepared by crystallization from the gas phase, following sublimation of a powder sample (ca. 12.5 mg) of the α polymorph. The sample was heated under vacuum inside a cylindrical curved-bottom vessel containing a cold finger (see Fig. S1a†). The cold finger was filled with acetone/dry ice at −78 °C (dry ice was added periodically to maintain the temperature at −78 °C). The temperature of the sample was increased from ambient temperature to a target temperature over a period of ca. 10 min, and was then left at the target temperature until sublimation was complete. During this time, microcrystalline powder samples deposited on the cold finger and on the outer glass tube (Fig. S1b†). The powder samples collected from these different parts of the apparatus were analyzed separately. In the initial experiments involving sublimation at 300 °C, the sample collected from the outer glass tube was found (from powder XRD analysis) to be a biphasic sample comprising primarily the β polymorph, but with a small amount of the α polymorph. In later experiments involving sublimation at 275 °C, the sample collected from the outer glass tube was found to be a monophasic sample of the β polymorph. In all experiments, the sample collected from the cold finger was a monophasic sample of the α polymorph.

Powder XRD

Powder XRD data were recorded at ambient temperature (21 °C) on a Bruker D8 Diffractometer (Ge-monochromated CuKα$_1$ radiation; Våntec detector covering 3° in 2θ; 2θ range, 4° to 70°; step size, 0.016°). The polycrystalline sample was packed into a glass capillary (0.7 mm diameter) which was flame sealed and placed on a foil-type sample holder. The data collection time was 64.5 h for the original biphasic sample and 191 h for the pure sample of the β polymorph. Although the powder XRD data were measured in the 2θ range from 4° to 70°, no peaks are observed below 2θ = 8° and no peaks are predicted for the β polymorph below 2θ = 8°. Thus, the region below 2θ = 8° was excluded from the powder XRD data used in Pawley fitting and Rietveld refinement.

To assess the degree of preferred orientation in the powder sample, two-dimensional powder XRD data were recorded at ambient temperature on an Agilent SuperNova diffractometer using CuKα radiation and an Atlas detector. The instrument

was configured with the two-dimensional detector perpendicular to the incident beam direction, with a sample-to-detector distance of 100 mm. The capillary containing the powder sample was oriented with the capillary axis perpendicular to the incident X-ray beam. Data were collected over a period of 10 s, with the capillary rotated through an angular range of 10° about the capillary axis during the data collection.

Electron diffraction

The dry powder of the β polymorph of ʟ-tyrosine was dispersed directly on a continuous carbon film supported by a 200-mesh copper grid, which was then mounted on a Gatan cryo-transfer tomography holder (No. 914). 3D-ED data were collected using a JEOL JEM-2100 transmission electron microscope (TEM) operating at an accelerating voltage of 200 kV, with the sample cooled to *ca.* 100 K. The 3D-ED data were recorded using the continuous rotation electron diffraction (cRED) method[51] by continuously tilting the goniometer with tilt speed 0.454° per second. During tilting, the crystal was tracked by sequential defocusing of the intermediate lens using the Instamatic software.[75] The data collection covered a rotation range of 98.41° (for more details, see Table S4†). The diffraction patterns were collected using the high-speed hybrid detection camera Timepix Quad (ASI) with integration time 0.5 s. The datasets were processed using X-ray Detector Software (XDS)[52] in order to extract intensity data. Structure refinement against 3D-ED data was carried out using SHELXL,[76] with isotropic atomic displacement parameters and atomic scattering factors for electrons. Hydrogen atoms were included within the structural model based on geometric restraints using the HFIX command in SHELXL.

Solid-state NMR spectroscopy

High-resolution solid-state ^{13}C NMR data were recorded at ambient temperature (20 °C) for samples of the α polymorph and the β polymorph of ʟ-tyrosine on a Bruker AVANCE III spectrometer at the U. K. High-Field (850 MHz) Solid-State NMR Facility (^{13}C Larmor frequency, 213.8 MHz; 4 mm HX probe; zirconia rotor; MAS frequency, 12 kHz) using ramped ^{1}H → ^{13}C cross-polarization with ^{1}H decoupling (using SPINAL-64) applied during acquisition. The total number of scans acquired was 32 for the α polymorph and 224 for the β polymorph, with a recycle delay of 60 s between each scan. The ^{13}C NMR data were referenced using ʟ-alanine, for which the carboxylate resonance was set to 177.9 ppm.

DFT-D calculations in conjunction with structure determination of the β polymorph from powder XRD data

Periodic DFT-D calculations were carried out at various stages of the structure determination of the β polymorph (particularly in conjunction with Rietveld refinement from powder XRD data) using the CASTEP program[77] (Academic Release Version 8.0). In particular, structural models were subjected to geometry optimization (with fixed unit cell) using ultrasoft pseudopotentials,[78] PBE functional,[79] semi-empirical dispersion correction using the TS correction scheme,[80] preserved space group

symmetry, periodic boundary conditions, a basis set cut-off energy of 700 eV and a Monkhorst–Pack grid[81] of minimum sample spacing $(0.05 \times 2\pi)$ Å^{-1}. Convergence criteria for geometry optimization were 0.01 eV Å^{-1} for the maximum atomic force, 0.00001 eV per atom on the total energy, and 0.001 Å for atomic displacements.

DFT-D calculations of solid-state ^{13}C NMR data

Solid-state ^{13}C NMR chemical shifts were calculated using CASTEP (Academic Release version 8.0) for the crystal structure of the β polymorph of ʟ-tyrosine obtained in the Rietveld refinement and following DFT-D geometry optimization (with fixed unit cell) of this structure, as described above. The same method was used to calculate the solid-state ^{13}C NMR chemical shifts for the published crystal structure[47] of the α polymorph of ʟ-tyrosine. In these calculations, the Gauge Including Projector Augmented Wave (GIPAW) approach[82–87] was used with a cut-off energy of 700 eV and PBE functional. A set of isotropic ^{13}C NMR shielding values was generated from the CASTEP calculations. From the isotropic ^{13}C NMR shielding value (σ_{calc}) calculated for each ^{13}C environment in the crystal structure, the corresponding calculated isotropic ^{13}C NMR chemical shift (δ_{calc}) was determined[84] from the equation $\delta_{calc} = \langle\delta_{exp}\rangle + \langle\sigma_{calc}\rangle - \sigma_{calc}$, where $\langle\delta_{exp}\rangle$ denotes the mean of the isotropic ^{13}C NMR chemical shifts determined from the experimental high-resolution solid-state ^{13}C NMR spectrum and $\langle\sigma_{calc}\rangle$ denotes the mean of the calculated isotropic ^{13}C NMR shielding values. For the β polymorph, $\langle\delta_{exp}\rangle = 116.30$ ppm and $\langle\sigma_{calc}\rangle = 54.21$ ppm. For the α polymorph, $\langle\delta_{exp}\rangle = 116.23$ ppm and $\langle\sigma_{calc}\rangle = 53.33$ ppm.

DFT-D calculations to compare the energetic properties of the α and β polymorphs of ʟ-tyrosine

After completing the Rietveld refinement, comparison of the energetic properties of the β polymorph and the previously reported α polymorph was carried out through DFT-D calculations using the FHI-aims software package[88] (date stamp: 191029). In this component of the work, different functionals and different methods for dispersion correction were considered, with FHI-aims selected for this study on account of its efficiency for periodic hybrid-DFT calculations.[89,90] Initially, the final refined crystal structure of the β polymorph from the Rietveld refinement and the previously published crystal structure of the α polymorph[47] were subjected to geometry optimization with fixed unit cell using the PBE-TS method, and based on a convergence criterion of 0.01 eV Å^{-1} for the maximum atomic force. Following the geometry optimization, single-point energy calculations were carried out using both generalized gradient approximation (GGA) and hybrid-GGA exchange-correlation functionals, specifically PBE[79] and PBE0,[91] coupled with either the Tkatchenko–Scheffler (TS) method[80] or the many-body dispersion (MBD) method[92] for dispersion correction. Thus, the complete set of exchange-correlation functionals considered in the single-point energy calculations were: PBE-TS, PBE-MBD, PBE0-TS and PBE0-MBD. All calculations were carried out with an "intermediate" basis set and relativistic effects were included *via* the scaled zeroth

order regular approximation.[88] A Γ-centred **k**-grid was used with a minimum sample spacing of $(0.05 \times 2\pi)$ Å^{-1} [testing with a denser **k**-grid sampling of $(0.04 \times 2\pi)$ Å^{-1} gave changes in relative energies of less than 1 meV]. The electronic structure self-consistent field (SCF) cycle was considered to be converged when changes in the electron density, the total energy and the sum of the eigenvalue energies were below 10^{-6} e a_0^{-3}, 10^{-6} eV and 10^{-6} eV, respectively.

Crystal structure prediction

Crystal structure prediction of L-tyrosine was carried out using the *ab initio* random structure searching (AIRSS) method,[45,46] which has recently been used successfully for other organic molecular crystals.[93,94] The AIRSS methodology for structure prediction involves two steps: (1) generation of trial structures by "intelligent" random searching, and (2) geometry optimization and energy-ranking of the trial structures using full periodic DFT calculations (PBE-TS).

Stage (1) considered crystal structures containing Z molecules per unit cell, with Z = 1, 2, 3, 4 or 6. Structures were generated from an initial, randomly placed molecule by applying the symmetry operators of all space groups. In this structure-generation step, the unit cell parameters were allowed to vary within constraints to give a volume per molecule within 10% of an initial estimate of 210 Å^3. Minimum intermolecular atom–atom distances were also specified in order to eliminate structures containing unreasonable intermolecular contacts. Searches with smaller unit cells (corresponding to Z = 1 or 2) were carried out with a global minimum separation of 2 Å. The minimum separations between intermolecular pairs of atomic species in the resulting low-energy structures were determined and used to specify minimum separations in searches with larger unit cells (corresponding to Z = 3, 4 or 6). In stage (2), "initial" geometry optimization (including relaxation of the unit cell) and energy ranking was achieved by DFT calculations using the CASTEP code,[77] which uses a plane-wave basis-set together with pseudopotentials to represent the core–valence interaction. All calculations used the PBEsol functional,[95] a plane-wave cut-off energy of 800 eV and a Brillouin zone sampling of $(0.07 \times 2\pi)$ Å^{-1}. The convergence criteria were set to the CASTEP default tolerances of 2×10^{-5} eV per atom on the total energy, 0.05 eV Å^{-1} on the maximum atomic force, 1×10^{-3} Å on the maximum atomic displacement, and 0.1 GPa on the maximum stress component. This initial structure prediction produced low-energy structures only for Z = 2 and Z = 4. Structures in non-chiral space groups were removed from the initial structure prediction list, giving only the chiral structures of interest. The relative energies of the predicted crystal structures generated by the two steps of the AIRSS methodology are shown in Table S1.†

Each of the predicted crystal structures resulting from the two stages of AIRSS was then examined by further DFT-D calculations using FHI-aims. First, a "precise" geometry optimization (including relaxation of unit cell parameters) was carried out on each structure using the PBE-TS method and based on a convergence criterion of 0.01 eV Å^{-1} for the

maximum atomic force. Second, for each structure obtained following the "precise" geometry optimization, a single-point calculation was carried out using the PBE0-MBD method to obtain a more reliable energy ranking.[72,73] The results of the energy ranking of the predicted crystal structures following the "precise" geometry optimization and the subsequent single-point PBE0-MBD calculations are given in Table S2.†

Comparison of crystal structures

All overlay plots for comparison of crystal structures (Fig. 2 and S5–S7†) were generated in Mathematica (Version 12.3).[96] In each case, the origins of the unit cells of the two structures were fixed at the same point, the *a*-axes of the two structures were aligned parallel to each other, and the *b*-axes of the two structures were aligned parallel to each other (we note that, for the monoclinic system, the *a*-axis and *b*-axis are orthogonal). For the overlay plots comparing structures with slightly different unit cells (Fig. 2 and S7†), both unit cells are shown (as indicated in the figure captions).

Data availability

Supporting Information contains: (i) additional figures discussed in the text, (ii) a detailed description of the crystal structure of the β polymorph of L-tyrosine and comparison to the α polymorph, (iii) tables of results from AIRSS calculations for crystal structure prediction of L-tyrosine, (iv) definition of *R*-factors used in direct-space structure solution from 3D-ED data and powder XRD data, (v) 3D-ED data statistics, (vi) computed values of isotropic ^{13}C NMR chemical shifts for the α and β polymorphs of L-tyrosine, (vii) crystallographic information (cif) file for the crystal structure of the β polymorph of L-tyrosine determined in the final Rietveld refinement from powder XRD data, and (viii) crystallographic information files for the five distinct crystal structures of L-tyrosine generated from the AIRSS structure prediction calculations (specifically, the predicted structure corresponding to the α polymorph, the predicted structure corresponding to the β polymorph, and the new predicted structures A, B and C).

Data and results from experimental and computational studies (with the exception of the DFT-D calculations using FHI-Aims) are available at: https://doi.org/10.17035/d.2022.0164412918. Results from DFT-D calculations using FHI-Aims are available at: https://doi.org/10.17172/NOMAD/2021.10.18-1.

The crystal structure of the β polymorph of L-tyrosine determined in the final Rietveld refinement from powder XRD data is also available at CCDC 2114085.

Author contributions

The research project was coordinated by K. D. M. H., with specific components of the research carried out as follows: preparation of the sample of the new polymorph of L-tyrosine (H. E. H., C. E. H.); measurement of powder XRD data (H. E. H., C. J. H. S.); measurement, analysis and interpretation of 3D-ED

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

References

1 L. B. McCusker, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1991, **47**, 297–343.

2 A. K. Cheetham and A. P. Wilkinson, *Angew. Chem., Int. Ed. Engl.*, 1992, **31**, 1557–1570.

3 P. Lightfoot, M. Tremayne, K. D. M. Harris and P. G. Bruce, *J. Chem. Soc., Chem. Commun.*, 1992, 1012–1013.

4 K. D. M. Harris, M. Tremayne, P. Lightfoot and P. G. Bruce, *J. Am. Chem. Soc.*, 1994, **116**, 3543–3547.

5 D. M. Poojary and A. Clearfield, *Acc. Chem. Res.*, 1997, **30**, 414–422.

6 H. Tsue, M. Horiguchi, R. Tamura, K. Fujii and H. Uekusa, *J. Synth. Org. Chem., Jpn.*, 2007, **65**, 1203–1212.

7 W. I. F. David and K. Shankland, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2008, **64**, 52–64.

8 K. D. M. Harris, *Top. Curr. Chem.*, 2012, **315**, 133–178.

9 J. Martí-Rujas, *Dalton Trans.*, 2020, **49**, 13897–13916.

10 R. Vincent and P. A. Midgley, *Ultramicroscopy*, 1994, **53**, 271–282.

11 U. Kolb, T. Gorelik, C. Kübel, M. T. Otten and D. Hubert, *Ultramicroscopy*, 2007, **107**, 507–513.

12 X. Zou and S. Hovmoller, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2008, **64**, 149–160.

13 P. A. Midgley and A. S. Eggeman, *IUCrJ*, 2015, **2**, 126–136.

14 Y. F. Yun, X. D. Zou, S. Hovmoller and W. Wan, *IUCrJ*, 2015, **2**, 267–282.

15 L. Palatinus, C. Antunes Corrêa, G. Steciuk, D. Jacob, P. Roussel, P. Boullay, M. Klementová, M. Gemmi, J. Kopeček, M. C. Domeneghetti, F. Cámara and V. Petříček, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2015, **71**, 740–751.

16 L. Palatinus, P. Brázda, P. Boullay, O. Perez, M. Klementová, S. Petit, V. Eigner, M. Zaarour and S. Mintova, *Science*, 2017, **355**, 166–169.

17 T. Gruene, J. T. C. Wennmacher, C. Zaubitzer, J. J. Holstein, J. Heidler, A. Fecteau-Lefebvre, S. De Carlo, E. Muller, K. N. Goldie, I. Regeni, T. Li, G. Santiso-Quinones, G. Steinfeld, S. Handschin, E. van Genderen, J. A. van Bokhoven, G. H. Clever and R. Pantelic, *Angew. Chem., Int. Ed.*, 2018, **57**, 16313–16317.

18 M. Gemmi, E. Mugnaioli, T. E. Gorelik, U. Kolb, L. Palatinus, P. Boullay, S. Hovmöller and J. P. Abrahams, *ACS Cent. Sci.*, 2019, **5**, 1315–1329.

19 P. Brázda, L. Palatinus and M. Babor, *Science*, 2019, **364**, 667–669.

20 G. R. Woollam, P. P. Das, E. Mugnaioli, I. Andrusenko, A. S. Galanis, J. van de Streek, S. Nicolopoulos, M. Gemmi and T. Wagner, *CrystEngComm*, 2020, **22**, 7490–7499.

21 Y. Wang, T. Yang, H. Xu, X. Zou and W. Wan, *J. Appl. Crystallogr.*, 2018, **51**, 1094–1101.

22 L. Smrčok, M. Brunelli, M. Boča and M. Kuchárík, *J. Appl. Crystallogr.*, 2008, **41**, 634–636.

23 E. E. Avila, A. J. Mora, G. E. Delgado, R. R. Contreras, L. Rincon, A. N. Fitch and M. Brunelli, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2009, **65**, 639–646.

24 A. J. Florence, J. Bardin, B. Johnston, N. Shankland, T. A. N. Griffin and K. Shankland, *Z. Kristallogr.*, 2009, **30**, 215–220.

25 J. van de Streek and M. A. Neumann, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2010, **66**, 544–558.

26 D. V. Dudenko, P. A. Williams, C. E. Hughes, O. N. Antzutkin, S. P. Velaga, S. P. Brown and K. D. M. Harris, *J. Phys. Chem. C*, 2013, **117**, 12258–12265.

27 G. N. M. Reddy, D. S. Cook, D. Iuga, R. I. Walton, A. Marsh and S. P. Brown, *Solid State Nucl. Magn. Reson.*, 2015, **65**, 41–48.

28 M. Sardo, S. M. Santos, A. A. Babaryk, C. López, I. Alkorta, J. Elguero, R. M. Claramunt and L. Mafra, *Solid State Nucl. Magn. Reson.*, 2015, **65**, 49–63.

29 F. Fischer, M. U. Schmidt, S. Greiser and F. Emmerling, *Acta Crystallogr., Sect. C: Struct. Chem.*, 2016, **72**, 217–224.

30 A. E. Watts, K. Maruyoshi, C. E. Hughes, S. P. Brown and K. D. M. Harris, *Cryst. Growth Des.*, 2016, **16**, 1798–1804.

31 S. D. Gumbert, M. Körbitzer, E. Alig, M. U. Schmidt, M. R. Chierotti, R. Gobetto, X. Li and J. van de Streek, *Dyes Pigm.*, 2016, **131**, 364–372.

32 C. E. Hughes, G. N. M. Reddy, S. Masiero, S. P. Brown, P. A. Williams and K. D. M. Harris, *Chem. Sci.*, 2017, **8**, 3971–3979.

33 A. S. Tatton, H. Blade, S. P. Brown, P. Hodgkinson, L. P. Hughes, S. O. Nilsson Lill and J. R. Yates, *Cryst. Growth Des.*, 2018, **18**, 3339–3351.

34 C. J. H. Smalley, A. J. Logsdail, C. E. Hughes, D. Iuga, M. T. Young and K. D. M. Harris, *Cryst. Growth Des.*, 2022, **22**, 524–534.

35 E. Courvoisier, P. A. Williams, G. K. Lim, C. E. Hughes and K. D. M. Harris, *Chem. Commun.*, 2012, **48**, 2761–2763.

36 C. H. Görbitz, K. W. Törnroos and G. M. Day, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2012, **68**, 549–557.

37 O. Al Rahal, C. E. Hughes, P. A. Williams, A. J. Logsdail, Y. Diskin-Posner and K. D. M. Harris, *Angew. Chem., Int. Ed.*, 2019, **58**, 18788–18792.

38 P. A. Williams, C. E. Hughes and K. D. M. Harris, *Angew. Chem., Int. Ed.*, 2015, **54**, 3973–3977.

39 A. Gavezzotti, *Acc. Chem. Res.*, 1994, **27**, 309–314.

40 G. M. Day, *Crystallogr. Rev.*, 2011, **17**, 3–52.

41 S. L. Price, *Acc. Chem. Res.*, 2009, **42**, 117–126.

42 A. R. Oganov, C. J. Pickard, Q. Zhu and R. J. Needs, *Nat. Rev. Mater.*, 2019, **4**, 331–338.

43 S. L. Price, *Chem. Soc. Rev.*, 2014, **43**, 2098–2111.

44 S. M. Woodley, G. M. Day and R. Catlow, *Philos. Trans. R. Soc., A*, 2020, **378**, 20190600.

45 C. J. Pickard and R. J. Needs, *Phys. Rev. Lett.*, 2006, **97**, 045504.

46 C. J. Pickard and R. J. Needs, *J. Phys.: Condens. Matter*, 2011, **23**, 053201.

47 A. Mostad, H. M. Nissen and C. Romming, *Acta Chem. Scand.*, 1972, **26**, 3819–3833.

48 M. N. Frey, T. F. Koetzle, M. S. Lehmann and W. C. Hamilton, *J. Chem. Phys.*, 1973, **58**, 2547–2556.

49 Z. Liu and C. Li, *Biophys. Chem.*, 2008, **138**, 115–119.

50 R. Shirley, *The CRYSFIRE 2002 System for Automatic Powder Indexing: User's Manual*, The Lattice Press, Guildford, Surrey, U.K., 2002.

51 W. Wan, J. Sun, J. Su, S. Hovmöller and X. Zou, *J. Appl. Crystallogr.*, 2013, **46**, 1863–1873.

52 W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 133–144.

53 K. D. M. Harris, R. L. Johnston, C. E. Hughes, Z. Zhou, S. Habershon, G. W. Turner, P. A. Williams, B. M. Kariuki and E. Y. Cheung, *EAGER – A Computer Program for Direct-Space Structure Solution from Diffraction Data*, Cardiff University.

54 B. M. Kariuki, H. Serrano-González, R. L. Johnston and K. D. M. Harris, *Chem. Phys. Lett.*, 1997, **280**, 189–195.

55 K. D. M. Harris, R. L. Johnston and S. Habershon, *Struct. Bonding*, 2004, **110**, 55–94.

56 K. D. M. Harris, S. Habershon, E. Y. Cheung and R. L. Johnston, *Z. Kristallogr.*, 2004, **219**, 838–846.

57 B. M. Kariuki, K. Psallidas, K. D. M. Harris, R. L. Johnston, R. W. Lancaster, S. E. Staniforth and S. M. Cooper, *Chem. Commun.*, 1999, 1677–1678.

58 E. Tedesco, G. W. Turner, K. D. M. Harris, R. L. Johnston and B. M. Kariuki, *Angew. Chem., Int. Ed.*, 2000, **39**, 4488–4491.

59 D. Albesa-Jové, B. M. Kariuki, S. J. Kitchin, L. Grice, E. Y. Cheung and K. D. M. Harris, *ChemPhysChem*, 2004, **5**, 414–418.

60 P. A. Williams, C. E. Hughes, G. K. Lim, B. M. Kariuki and K. D. M. Harris, *Cryst. Growth Des.*, 2012, **12**, 3104–3113.

61 X. Ma, G. K. Lim, K. D. M. Harris, D. C. Apperley, P. N. Horton, M. B. Hursthouse and S. L. James, *Cryst. Growth Des.*, 2012, **12**, 5869–5872.

62 J. Martí-Rujas, L. Meazza, G. K. Lim, G. Terraneo, T. Pilati, K. D. M. Harris, P. Metrangolo and G. Resnati, *Angew. Chem., Int. Ed.*, 2013, **52**, 13444–13448.

63 I. Brekalo, W. Yuan, C. Mottillo, Y. Lu, Y. Zhang, J. Casaban, K. T. Holman, S. L. James, F. Duarte, P. A. Williams, K. D. M. Harris and T. Friščić, *Chem. Sci.*, 2020, **11**, 2141–2147.

64 T. Sun, C. E. Hughes, L. Guo, L. Wei, K. D. M. Harris, Y. Zhang and Y. Ma, *Angew. Chem., Int. Ed.*, 2020, **59**, 22638–22644.

65 I. J. Bruno, J. C. Cole, M. Kessler, J. Luo, W. D. S. Motherwell, L. H. Purkis, B. R. Smith, R. Taylor, R. I. Cooper, S. E. Harris and A. G. Orpen, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 2133–2144.

66 F. H. Allen, O. Kennard, D. G. Watson, L. Brammer, A. G. Orpen and R. Taylor, *J. Chem. Soc., Perkin Trans. 2*, 1987, S1–S19.

67 H. M. Rietveld, *J. Appl. Crystallogr.*, 1969, **2**, 65–71.

68 A. Coelho, *TOPAS Academic, Version 5*, 2012.

69 G. S. Pawley, *J. Appl. Crystallogr.*, 1981, **14**, 357–361.

70 A. March, *Z. Kristallogr.*, 1932, **81**, 285–297.

71 W. A. Dollase, *J. Appl. Crystallogr.*, 1986, **19**, 267–272.

72 A. G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkatchenko, A. R. Oganov, A. L. Rohl, M. E. Tuckerman and B. Kahr, *Chem. Sci.*, 2017, **8**, 4926–4940.

73 J. Hoja, H. Y. Ko, M. A. Neumann, R. Car, R. A. DiStasio and A. Tkatchenko, *Sci. Adv.*, 2019, **5**, eaau3338.

74 M. Nyman and G. M. Day, *CrystEngComm*, 2015, **17**, 5154.

75 M. O. Cichocka, J. Ångström, B. Wang, X. Zou and S. Smeets, *J. Appl. Crystallogr.*, 2018, **51**, 1652–1661.

76 G. M. Sheldrick, *Acta Crystallogr., Sect. C: Struct. Chem.*, 2015, **71**, 3–8.

77 S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson and M. C. Payne, *Z. Kristallogr.*, 2005, **220**, 567–570.

78 D. Vanderbilt, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1990, **41**, 7892–7895.

79 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

80 A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.

81 H. J. Monkhorst and J. D. Pack, *Phys. Rev. B: Solid State*, 1976, **13**, 5188–5192.

82 C. J. Pickard and F. Mauri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2001, **63**, 245101.

83 J. R. Yates, C. J. Pickard and F. Mauri, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2007, **76**, 024401.

84 R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates and V. Zorin, *Magn. Reson. Chem.*, 2007, **45**, S174–S186.

85 T. Charpentier, *Solid State Nucl. Magn. Reson.*, 2011, **40**, 1–20.

86 C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azaïs, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri and C. J. Pickard, *Chem. Rev.*, 2012, **112**, 5733–5779.

87 S. E. Ashbrook and D. McKay, *Chem. Commun.*, 2016, **52**, 7186–7204.

88 V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput. Phys. Commun.*, 2009, **180**, 2175–2196.

89 X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter and M. Scheffler, *New J. Phys.*, 2012, **14**, 053020.

90 S. V. Levchenko, X. Ren, J. Wieferink, R. Johanni, P. Rinke, V. Blum and M. Scheffler, *Comput. Phys. Commun.*, 2015, **192**, 60–69.

91 C. Adamo and V. Barone, *J. Chem. Phys.*, 1999, **110**, 6158–6170.

92 A. Ambrosetti, A. M. Reilly, R. A. DiStasio and A. Tkatchenko, *J. Chem. Phys.*, 2014, **140**, 18A508.

93 M. Zilka, D. V. Dudenko, C. E. Hughes, P. A. Williams, S. Sturniolo, W. T. Franks, C. J. Pickard, J. R. Yates, K. D. M. Harris and S. P. Brown, *Phys. Chem. Chem. Phys.*, 2017, **19**, 25949–25960.

94 R. Mathew, K. A. Uchman, L. Gkoura, C. J. Pickard and M. Baias, *Magn. Reson. Chem.*, 2020, **58**, 1018–1025.

95 Ł. Szeleszczuk, D. M. Pisklak and M. Zielińska-Pisklak, *J. Comput. Chem.*, 2018, **39**, 1300–1306.

96 Wolfram Research, Inc., *Mathematica 12.3*, Champaign, Illinois, 2021.