










ARTICLE


<https://doi.org/10.1038/s41467-021-23208-7>

OPEN

Structure determination of an amorphous drug through large-scale NMR predictions

Manuel Cordova ^{1,2}, Martins Balodis¹, Albert Hofstetter ¹, Federico Paruzzo¹, Sten O. Nilsson Lill³, Emma S. E. Eriksson ³, Pierrick Berruyer ¹, Bruno Simões de Almeida¹, Michael J. Quayle ⁴, Stefan T. Norberg ⁵, Anna Svensk Ankarberg ⁵, Staffan Schantz ⁵✉ & Lyndon Emsley ^{1,2}✉

Knowledge of the structure of amorphous solids can direct, for example, the optimization of pharmaceutical formulations, but atomic-level structure determination in amorphous molecular solids has so far not been possible. Solid-state nuclear magnetic resonance (NMR) is among the most popular methods to characterize amorphous materials, and molecular dynamics (MD) simulations can help describe the structure of disordered materials. However, directly relating MD to NMR experiments in molecular solids has been out of reach until now because of the large size of these simulations. Here, using a machine learning model of chemical shifts, we determine the atomic-level structure of the hydrated amorphous drug AZD5718 by combining dynamic nuclear polarization-enhanced solid-state NMR experiments with predicted chemical shifts for MD simulations of large systems. From these amorphous structures we then identify H-bonding motifs and relate them to local intermolecular complex formation energies.

¹Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. ²National Centre for Computational Design and Discovery of Novel Materials MARVEL, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. ³Early Product Development and Manufacturing, Pharmaceutical Sciences, R&D, AstraZeneca, Gothenburg, Sweden. ⁴New Modalities and Parenteral Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, Gothenburg, Sweden. ⁵Oral Product Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, Gothenburg, Sweden. ✉email: Staffan.Schantz@astrazeneca.com; lyndon.emsley@epfl.ch

Determination of the structure of organic solids is a key step in many areas of chemistry, and in particular for the design of efficient and safe pharmaceutical compounds. This requires methods for atomic-level structure determination. Whereas X-Ray diffraction (XRD) is the established gold standard when single crystals are available, atomic-level structure determination is much more challenging in powder samples, and even more so if the structures are disordered. Indeed, while there has been much progress towards complete structure determination in crystalline powders, by powder XRD^{1,2} or particularly from solid-state nuclear magnetic resonance (NMR) approaches^{3–5}, the disorder inherent to amorphous solids makes structure determination elusive.

For example, the structure, accessible surface area, and stability of amorphous drug formulations are of high current interest^{6–11}, particularly since the bioavailability or dissolution rate of poorly soluble compounds in crystalline forms is often a severe limitation on the chemical space available for development of active pharmaceutical ingredients (APIs), and since the uptake of poorly soluble drugs can be significantly enhanced in amorphous formulations. In particular hydrated amorphous phases, wherein water molecules closely interact with the drug through hydrogen bonds, have been investigated for several systems of pharmaceutical interest^{12–14}.

Solid-state NMR is among the most popular methods to study the structure of amorphous materials. While two-dimensional correlation experiments are able to identify intermolecular contacts between atom pairs^{15–17}, obtaining complete atomic-level structures is a challenge due to the disordered molecular environments present in amorphous solids.

Comparing chemical shifts obtained from solid-state NMR experiments with shifts computed using density functional theory (DFT) methods for trial structures generated using crystal structure prediction (CSP) or molecular dynamics (MD) approaches has proven able to determine the structures of powdered crystalline organic (and hybrid) solids^{3,18,19}. One limitation of this procedure lies in the accuracy achievable by the DFT method used. While plane-wave gauge-including projector augmented wave (GIPAW) is the most popular approach to computing chemical shifts in crystalline solids^{20,21}, the increased computational cost incurred by the use of a plane-wave basis set instead of an atomic orbital basis set typically limits the level of theory to generalized gradient approximation (GGA) functionals. An alternative fragment- and cluster-based approach involving atomic orbital basis sets has recently been developed by Hartman et al. and allows the use of hybrid functionals, increasing the accuracy of computed chemical shifts at a computational cost comparable to GIPAW^{22–25}.

Candidate structures of amorphous compounds can be generated using MD simulations^{26–28}, however the large cell sizes required to model amorphous structures, together with the large number of structures needed to describe amorphous ensembles, have so far prevented the computation of chemical shifts in these systems. Recently a new approach to predicting chemical shifts in molecular solids was introduced²⁹ by training a machine-learning model on GIPAW DFT calculated chemical shifts for 3546 diverse structures from the Cambridge structural database (CSD)³⁰. This model, dubbed ShiftML, can now be used to predict chemical shifts for any molecular solid (currently limited to solids comprised of C, H, N, O, S atoms) in a matter of seconds^{29,31}, which opens up the possibility for large-scale shift computations on large structures.

Here, we investigate the structure of anhydrous crystalline AZD5718 (**1**) form A, by combining measured ¹H, ¹³C and ¹⁵N chemical shifts obtained using DNP-enhanced NMR experiments from a powder sample, CSP, and DFT chemical shift

computations. The structure is validated with that obtained from single-crystal XRD. We then model the hydrated amorphous drug with different water contents using MD simulations and obtain predicted NMR spectra for large structural ensembles using machine learned chemical shifts. We then analyse the ensembles to identify the different hydrogen bonding motifs present in the amorphous structures, by comparison of the experimental and predicted chemical shift distributions associated with these structural motifs. From the amorphous structures we also compute the interaction energy between molecules of **1** and their environment, and we relate the energies to the local hydrogen bonding motifs.

Results

NMR crystallography. No polymorphism was observed for anhydrous crystalline **1**, nor were crystalline hydrates identified. The crystal structure of anhydrous crystalline **1** Form A was determined using a chemical shift-based NMR crystallographic approach. This involves the combination of the assigned experimental chemical shifts with CSP and computed chemical shifts. The ¹H, ¹³C and ¹⁵N resonances of AZD5718 (Fig. 1e) were assigned using one-dimensional proton, carbon and nitrogen MAS NMR experiments (Fig. 1a–c), as well as two-dimensional refocused ¹³C–¹³C INADEQUATE and ¹H–¹³C HETCOR experiments (Fig. 1d–e) as detailed in Supplementary Discussion. A set of DFT-D optimized candidate structures was generated using an internally developed rapid CSP approach, then the assigned experimental chemical shifts were compared to the shifts computed using the cluster- and fragment-based DFT approach introduced by Hartman et al.^{22–24} for each structure in order to determine the experimental structure from the set of candidates. The structure determined using single-crystal X-ray diffraction was included and compared to the CSP set. The lowest energy CSP candidate (structure #1) was found to be structurally similar to the X-ray structure (as discussed in Supplementary Discussion). The comparisons of the experimental and computed ¹H and ¹³C chemical shifts are shown in Fig. 2a, b, respectively. The root-mean-square errors (RMSEs) obtained for ¹H suggested that structure #1 best matches the experiment, while ¹³C chemical shift results identified the X-ray structure as the best match. In addition, the DFT-D energy per molecule of structure #1 was found to be the lowest among the CSP set (x-axis in Fig. 2a–b). This also indicates that the force field used for the CSP procedure accurately describes the crystalline system, and supports the identification of candidate #1 as being the crystal structure. In order to elucidate the ambiguity between candidate #1 and the XRD structure, and to obtain a quantitative comparison of all candidates, a Bayesian probabilistic analysis was carried out using the approach introduced by Engel et al.³¹. The two main advantages of using this method to determine the structure that best matches experiment are the quantitative determination of the confidence in the identification of the experimental structure on a continuous scale from 0 to 100%, and the combined use of NMR results for several elements, increasing the accuracy of the identification.

Figure 2c shows the results obtained with the Bayesian approach, represented as a principal component analysis (PCA) plot. This plot is a two-dimensional representation of the similarity of the different candidate structures according to their computed chemical shifts. The computed Bayesian probability of each structure to be the experimental crystal structure is represented by the area of the blue disk around each point. Using both ¹H and ¹³C chemical shifts, candidate #1 is found to be the most probable crystal structure, with 99.7% confidence. Although the structure determined by X-ray diffraction (labelled

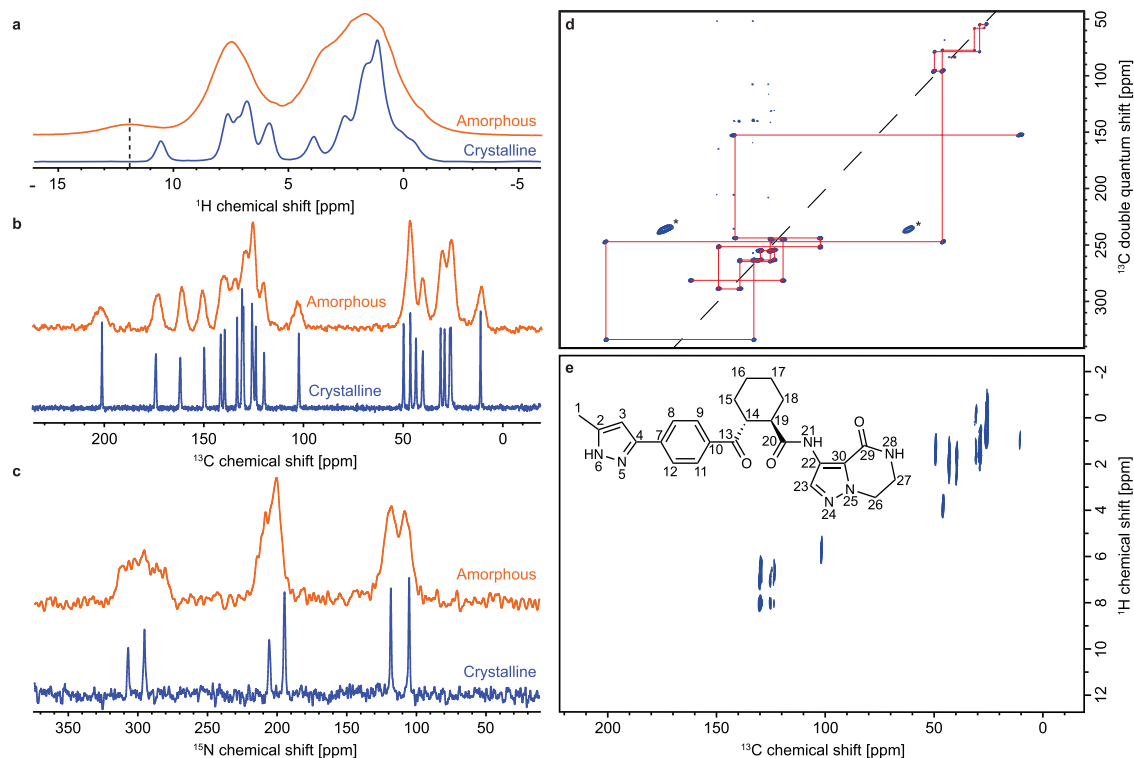


Fig. 1 Solid-state NMR experiments. **a**, ^1H , **b** ^{13}C and **c** ^{15}N MAS NMR spectra of crystalline (blue) and amorphous (orange) **1**. **d** ^{13}C - ^{13}C DNP enhanced solvent suppressed INADEQUATE and **e**, ^1H - ^{13}C HETCOR spectra of crystalline **1**. The dashed black line in (**a**) indicates the chemical shift assigned to the proton bound to N6 in the amorphous sample. In (**d**), the ^{13}C peaks denoted by a star at 60 and 170 ppm are attributed to impurities introduced during the NMR sample preparation. The chemical structure and labelling scheme of **1** is shown in (**e**).

XRD) appears closer to the experimental results (red cross) in the first two chemical shift principal components in Fig. 2c, including the complete chemical shift space identifies candidate #1 as the structure that best matches experiment, as indicated by its associated confidence. Figure 2c highlights the similarity of the selected structures in terms of their chemical shifts, in the two dimensions that display the largest variance.

Comparison of the structures determined via XRD and NMR crystallography yielded a RMSD₁₅ (root mean square deviation of the atomic positions in 15 molecules, ignoring hydrogen positions) of 0.42 Å. The main difference between the two structures lies in the conformation of the bicyclo ring (see Supplementary Fig. 2). Single-molecule heavy atom RMSD was found to be 0.22 Å, and was decreased to 0.15 Å after omitting the two carbons of the bicyclo ring (labelled 26 and 27 in Fig. 1e).

Unlike X-ray diffraction, NMR is highly sensitive to hydrogen nuclei, making it the method of choice for validating the tautomeric form of **1**. Indeed, any of the two nitrogen atoms of the pyrazole ring (labelled 5 and 6 in Fig. 1e) can be protonated in the crystalline sample. After computing ^1H , ^{13}C and ^{15}N shifts for the two possible tautomers displayed in Fig. 2d and comparing them with the experimental shifts (Fig. 2e), the resulting chemical shift RMSE was found to be consistently lower for tautomer A, by a factor of 1.3 for ^1H , 2.6 for ^{13}C and 8.1 for ^{15}N . This unambiguously identifies tautomer A as the crystal structure. The position of the N-H proton on the pyrazole ring is crucial in setting up amorphous structures able to describe the properties of the amorphous phase of **1**.

The atomic displacement parameter (ADP) tensors of all atoms in the structure determined by NMR crystallography were obtained as described in ref. 32. Simulation details are given in Supplementary Methods. Figure 2f shows the ORTEP plot of the ADP tensors corresponding to a ^1H chemical shift RMSE of 0.34 ppm. This value corresponds to the estimated error of ^1H

chemical shifts computed with the fragment- and cluster-based approach²⁴. The average value of the ADPs is 0.00025 Å².

Hydrogen bonding motifs in the amorphous phase. Knowledge of the structure of AZD5718 in the amorphous phase is key to understanding its physicochemical properties. Investigation of the amorphous structure of **1** was performed using NMR experiments combined with ShiftML predicted chemical shifts for MD ensembles.

Comparison of the proton, carbon and nitrogen NMR spectra in crystalline and amorphous **1** shown in Fig. 1a–c displays the overall broadening of the NMR signal typical of amorphous compounds. Apart from this observation, the chemical shifts do not display a significant change between the two phases of the compound. This suggests that **1** does not undergo large amplitude structural rearrangements upon transition from the crystalline to the amorphous state. The main difference between the NMR spectra of the two phases of **1** lies in the displacement of the ^1H resonance corresponding to the proton attached to the nitrogen labelled 6 (see Fig. 1e) from 10.6 ppm in the crystalline sample to 11.8 ppm in the amorphous form. This suggests a change in the hydrogen bonding network in the structure³³.

To better understand the structural differences between the amorphous and crystalline phases of **1**, we generated MD models of the amorphous structure at different hydration levels ranging from 0% to 4% (w/w) water content. This range of water content is representative of the experimental water content under real conditions, as confirmed by dynamic vapour sorption. Considering the large size of the simulation cells (128 molecules of **1** and up to 132 water molecules) and the large number of structures generated by MD, DFT computation of chemical shifts in these model systems would not be feasible. The machine learning model ShiftML was thus used to predict chemical shifts in these

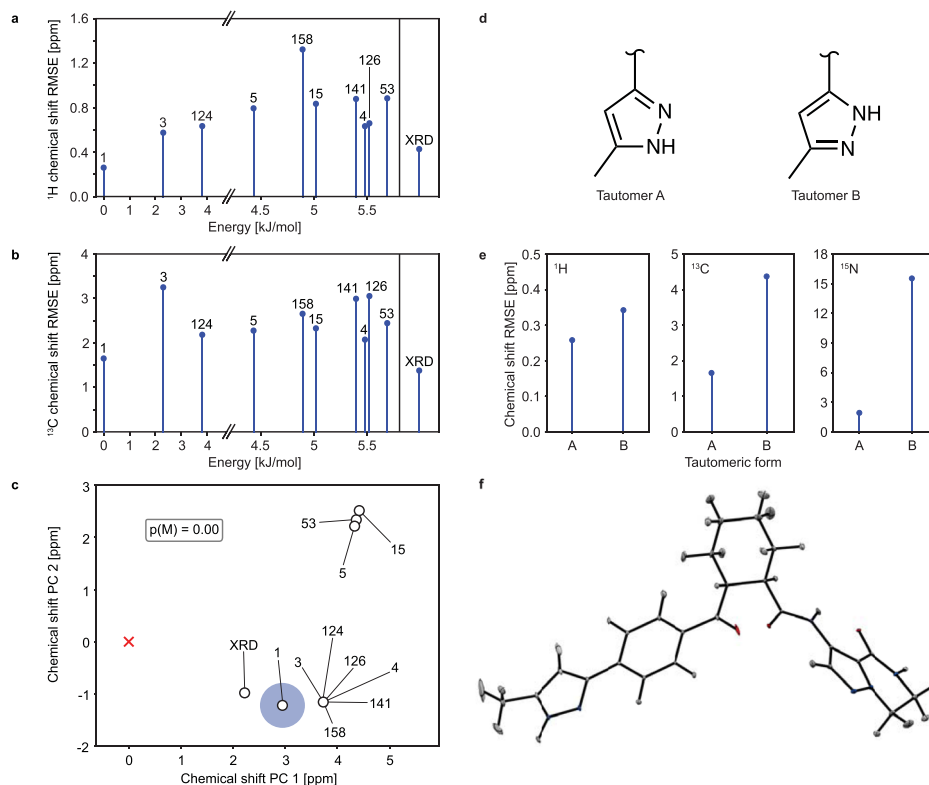


Fig. 2 NMR crystallography of AZD5718. **a** ^1H and **b** ^{13}C chemical shift RMSEs of the ten lowest DFT-D energy CSP candidates and of the crystal structure determined by X-ray diffraction (labelled XRD). The label rank in terms of increasing force-field energy per molecule of (**1**) of each candidate is indicated above each point. **c** Two-dimensional projection of the similarity of the computed ^1H and ^{13}C chemical shifts of the candidate structures to the experimental data (red cross). The probability of each candidate to match experiment is represented by the area of the blue disk. $p(M)$ denotes the probability that a virtual candidate, which represents structures potentially missing from the CSP candidate pool, matches experiment. The virtual candidate is represented by the mean of the shifts of all CSP candidates, and a high probability $p(M)$ suggests that the CSP candidate pool may not include the experimentally observed structure. **d** Chemical structures of the two tautomers of **1** considered, labelled as A and B. **e** Agreement between ^1H , ^{13}C and ^{15}N experimental and DFT computed chemical shifts for the two tautomers. **f** ORTEP plot of the ADP tensors for the NMR structure of **1** drawn at the 90% probability level.

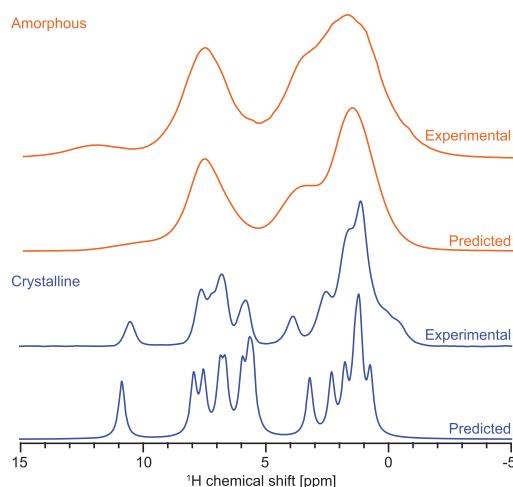


Fig. 3 Predicted spectra. Predicted and experimental ^1H NMR spectra of crystalline (blue) and amorphous (orange) **1**. The predicted spectrum of amorphous **1** was obtained by considering only the 4% w/w water MD simulations.

structures. The predicted spectra obtained for the crystalline structure and obtained by summing the spectra from 202 full cell snapshots of the 4% water MD simulations (i.e., 25,856 molecules of **1** and 26,664 water molecules) are displayed in Fig. 3. The ^1H

chemical shift RMSE obtained by comparing shifts predicted by ShiftML from the crystal structure with the experiment was found to be 0.61 ppm.

Although no clear peak is observed at 11.8 ppm for the amorphous structure, the population of predicted shifts above 11 ppm was found to increase slightly with increasing water content (see Supplementary Fig. 3). This behaviour suggests that interaction of **1** with water molecules does promote deshielding of the proton attached to the nitrogen labelled 6.

The predicted chemical shifts obtained were related to structural motifs in the model amorphous structures by identifying the different hydrogen bonding patterns (where the criteria are given in the methods section below) present in the structures and associated predicted ^1H shifts of the hydrogen bond donor groups. Figure 4a displays the chemical shift distributions of the most often occurring hydrogen bonding motifs. The atom most commonly bound to the N6-H group was found to be O20 (where O20 is the oxygen bound to C20), which corresponds to the hydrogen bond found in crystalline AZD5718. This is an indication that the structure of the amorphous compound is broadly similar to that of its crystalline counterpart. Over all analysed simulation snapshots (corresponding to an average water content of 1.16% (w/w), or about 3.4 times more molecules of **1** than water), water was found to be the fourth most occurring hydrogen bonding partner to the N6-H group, and was found to lead to the most pronounced deshielding of the hydrogen bond donor proton (motif D in Fig. 4b).

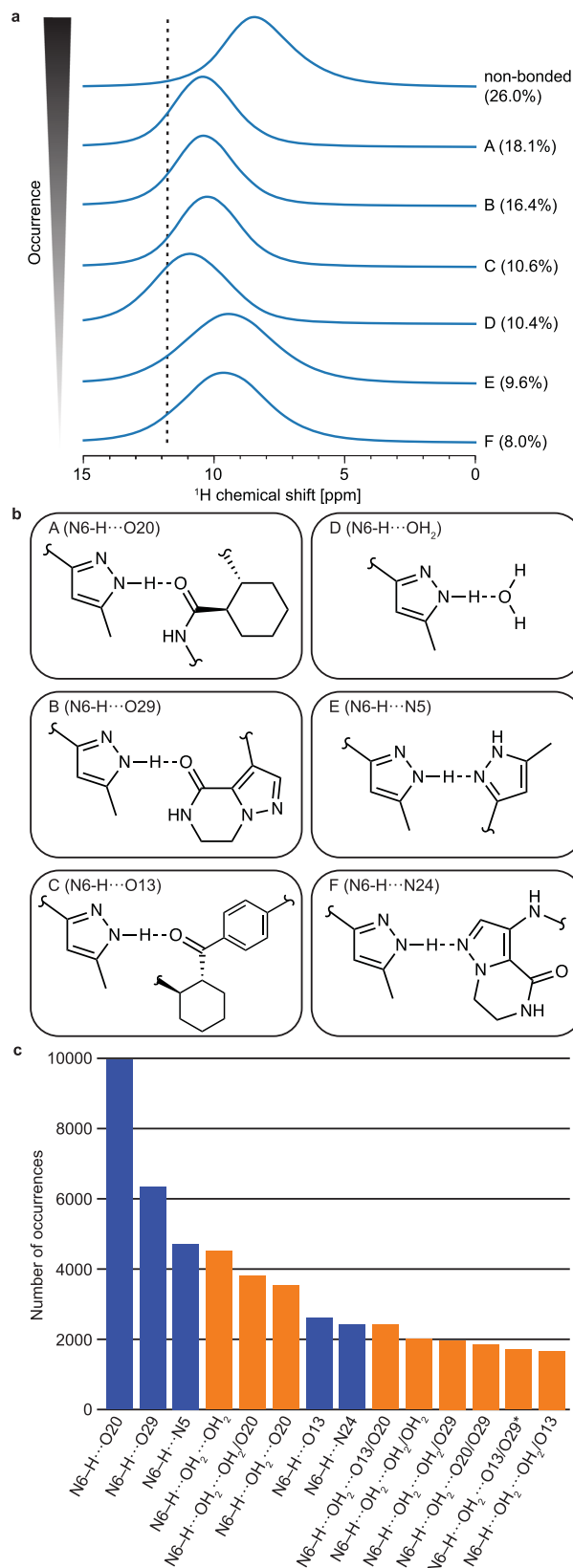


Fig. 4 H-bonding motifs in amorphous AZD5718. **a** Predicted spectra obtained using the predicted ^1H chemical shifts of the most often occurring N-H bonding motifs involving N6 for 11 evenly spaced snapshots of all amorphous simulations of each water content. The percentages next to the spectra denote the fraction of bonding motifs their corresponding pattern represents, including the instances where no H-bonded neighbour was identified. The dashed vertical line indicates the experimental shift observed in amorphous **1** and assigned to the proton bound to N6.

b Hydrogen bonding motifs associated with the spectra in **(a)**. **c** Number of occurrences of extended H-bonding motifs yielding a predicted chemical shift above 11 ppm for every snapshot of the 4% water simulations. Only the patterns corresponding to the top 75% of all shifts above 11 ppm were selected. The orange bars represent the bonding motifs involving water, and the blue ones correspond to the motifs that do not involve water. Two secondary neighbours from the same molecule are indicated by an asterisk. In **(b)** and **(c)**, *On* indicates the oxygen atom bonded to carbon *n*.

N6-H...OH₂ motifs yielding a chemical shift above 11 ppm in all snapshots of the amorphous 4% water MD simulations, and obtained the secondary neighbours, bonded to the water protons. We restricted this analysis to the simulations with the highest water content, as bonding of water was found to lead to the largest deshielding of the proton attached to N6 (see Fig. 4a), and as a larger number of water molecules in the simulation promotes extended hydrogen bonding motifs. Figure 4c shows the occurrences of extended hydrogen bonding patterns involving water, as well as the motifs made of pairs of H-bonded molecules of **1**, yielding a predicted shift above 11 ppm. The most often occurring pattern is the hydrogen bond present in the crystalline phase of the compound (N6-H...O20). When the H-bonded molecule is water, secondary neighbours are often found to be other water molecules, suggesting the formation of small clusters of water between AZD5718 molecules. Superpositions of ten instances of two hydrogen bonding motifs, N6-H...O20 and N6-H...OH₂...OH₂/O20, are shown in Fig. 5. It was observed that molecules of **1** being secondary H-bonded neighbours of N6-H generally lie away from the molecule bearing the hydrogen bond donor, indicating that steric clashes may constrain the possible geometries of hydrogen bonding in the amorphous form.

Formation energies of intermolecular complexes. Obtaining the formation energies of the supramolecular complexes of molecules of **1** and their surroundings can help determine which hydrogen bonding pairs lead to overall more favourable intra- and intermolecular interactions. After computing the formation energies, including the conformational energy of the probe molecule, using the semiempirical DFTB3-D3H5 method, the results were gathered as a function of the hydrogen bonding motifs in which the probe molecule was involved as the hydrogen bond donor. The relative formation energy was defined as the difference between the formation energy of each instance and the mean formation energy of all instances where no hydrogen bond acceptor was found for the selected hydrogen bond donor.

Figure 6 shows the relative formation energy for each hydrogen bond donor and acceptor identified. Bonding of any N-H group to water was found to yield the most favourable interactions. Over all the simulation snapshots analysed for hydrogen bonding motifs, 6.1% of N21-H chemical groups were found to form an intramolecular hydrogen bond with the carbonyl labelled 29. This number may however be underestimated, as the same intramolecular hydrogen bond is found in the crystal structure, with a bond angle of 128.8°. This hydrogen bond would thus not be identified using the cutoff values selected here.

Because a single water molecule can form two hydrogen bonds involving its hydrogen atoms and two additional ones involving its oxygen atom, more extended hydrogen bonding motifs are likely to be observed for AZD5718 molecules bound to water. In order to investigate these extended patterns, we extracted

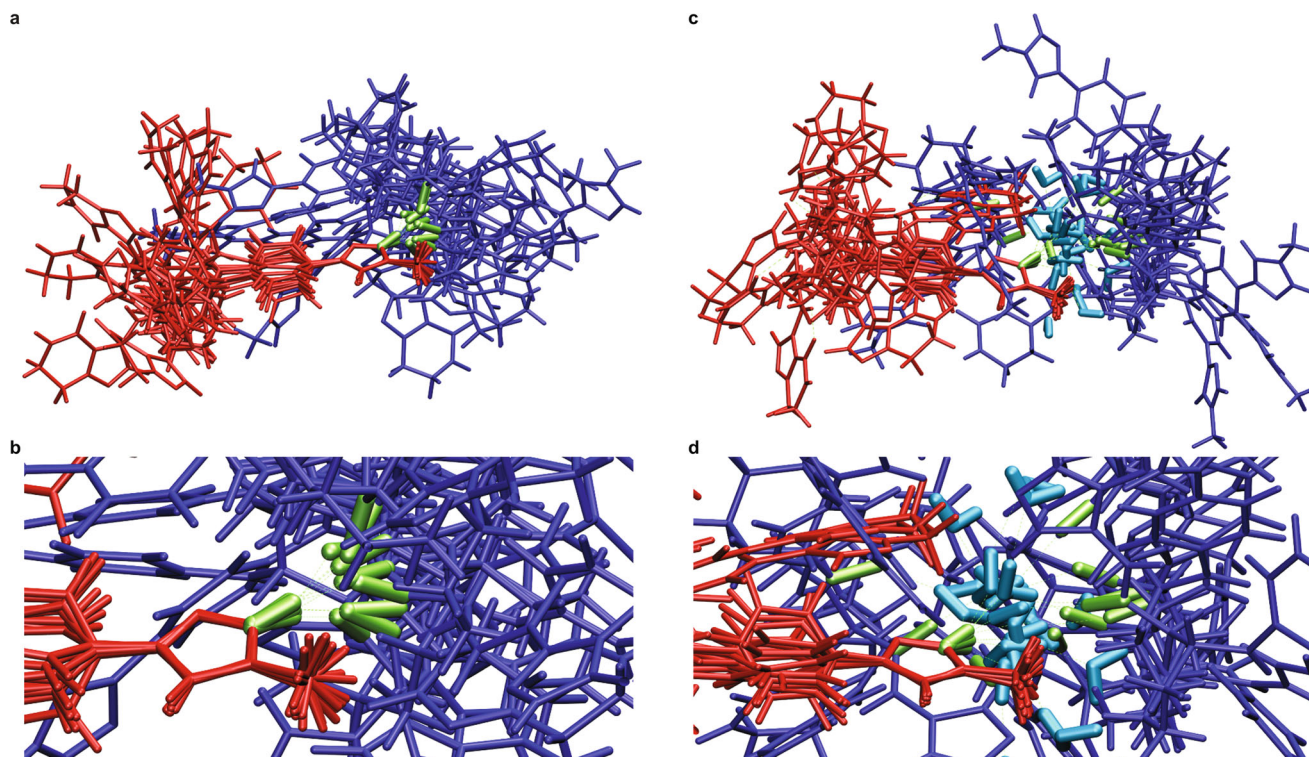


Fig. 5 Complete structures and H-bonding motifs. **a** Superposition of 10 instances of the N6-H...O2O bonding motif. **b** close-up view of the hydrogen bonding region in **(a)**. **c** Superposition of 10 instances of the N6-H...OH₂...OH₂/O2O bonding motif. **d** close-up view of the hydrogen bonding region in **(c)**. The red molecule represents **1** bearing the hydrogen bond donor (N6-H), the dark blue molecule represents **1** bearing the hydrogen bond acceptor, water molecules are coloured in cyan and the atoms of **1** involved in the hydrogen bonding motif are coloured in green.

Discussion

The atomic-level structure and hydrogen bonding patterns of the hydrated amorphous phase of AZD5718 were determined through solid-state NMR chemical shifts, MD simulations with various water contents, and machine learned chemical shifts. The chemical shifts associated with possible hydrogen bonding motifs generated from MD simulations were compared to experimental NMR spectra in order to identify the most commonly occurring intermolecular interactions in the amorphous material. Bonding of N6-H to water was found to yield the largest deshielding of the proton involved in the hydrogen bond, and best described the experimental shift observed in the amorphous sample. This intermolecular bond to water was also associated with more favourable intermolecular complex formation energies as compared to direct H-bonding between two AZD5718 molecules. These favourable water-AZD5718 interactions highlight the potential ability of water to prevent physical aging of the amorphous drug.

The combination of the three techniques presented here was crucial in elucidating the structure of this amorphous material through a large scale direct comparison of experimental chemical shifts with predicted shifts from MD structures. While solid-state NMR has already been used in tandem with MD simulations of amorphous materials, previous work have generally used molecular dynamics either to relate relative NMR peak areas to statistical ratios of different types of interactions^{28,34}, or to generate conformational ensembles from which small supramolecular clusters are extracted for DFT shift computation³⁵. Overall, the method presented here can be applied to a wide range of disordered organic systems to determine their complete atomic-level structures from their NMR spectra.

The structure of the crystalline form was also determined using NMR crystallography to within a positional error of 0.1 Å and

was confirmed to be almost identical to the structure obtained with single crystal X-Ray diffraction.

Methods

NMR experiments. Both crystalline and amorphous forms of **1** were provided by AstraZeneca. The samples were stored at equilibrium with the environment at ~22 °C and 20% relative humidity prior to NMR analysis. The room temperature NMR experiments were performed on Bruker Ascend 500 wide-bore Avance III, Bruker 800 Ultrashield plus narrow-bore and 900 US² wide-bore Avance Neo NMR spectrometers. DNP solid-state NMR spectroscopy experiments were performed on a 400 MHz Avance III HD Bruker spectrometer. The spectrometer is equipped with a low temperature magic angle spinning (LTMAS) 3.2 mm probe and connected through a corrugated waveguide to a 263 GHz gyrotron capable of outputting ca. 5–10 W of continuous wave microwaves. All chemical shifts were referenced to alanine. For more details including experimental setup and the sample preparation see Supplementary Methods.

NMR crystallography. The candidate crystal structures were generated using a Monte-Carlo parallel tempering method³⁶ followed by lattice energy minimisation using an internally developed force-field. The 190 most stable candidates were selected for full DFT-D optimization at the PBE level of theory. Chemical shifts for the ten lowest energy candidates were computed at the PBE0 level of theory using the fragment- and cluster-based approach developed by Hartman et al.^{22–24}. The conversion from isotropic shielding to chemical shift was performed by linear regression between the obtained shieldings and experimental isotropic chemical shifts. The analysis of the positional uncertainty of the crystal structure was performed as described by Hofstetter et al.³² by computing shifts of perturbed crystal structures obtained through low-temperature MD simulations of candidate #1 and relating chemical shift deviations to positional deviations. Chemical shift computations were also performed on an extended set of the 81 following lowest energy candidates, but did not lead to lower shift RMSEs than candidate #1. Further computational details are given in the Supplementary Methods.

Molecular dynamics simulation of amorphous structures. The amorphous structure of AZD5718 was modelled by carrying out MD simulations on periodic amorphous cells containing 128 molecules of **1** and a variable number of water molecules. Five cells of each water content; 0, 0.5, 1.0 and 2.0% (w/w, 0, 16, 32 and 65 water molecules in each cell, respectively), and two cells of 4% water (w/w, 132

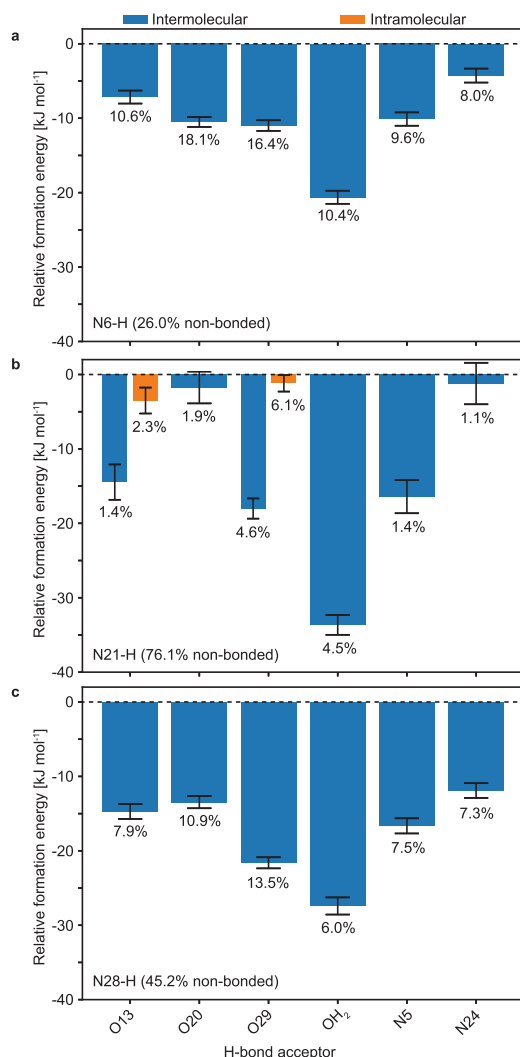


Fig. 6 Relative formation energy of the hydrogen bonding motifs. Mean computed formation energies of intermolecular complexes for the H-bond acceptor connected to **a** N6-H, **b** N21-H or **c** N28-H of the probe molecule. The percentage under each bar indicates the fraction of the N-H group bonded to the corresponding H-bond acceptor. Only the H-bond acceptors making up at least 1% of all instances analysed are displayed. The error bars indicate the standard error of the mean of the relative formation energies.

water molecules in each cell) were generated. After equilibration for 1 ns using the canonical NVT ensemble at 298 K followed by 10 ns using the isothermal-isobaric ensemble (NPT) at 298 K and 1 bar, production simulations were carried out for 600 ns using the NPT ensemble at 298 K and 1 bar. Models of the amorphous structure were obtained by extracting 1001 evenly spaced snapshots from the last 100 ns of each MD simulation, corresponding to 100 ps time steps between the extracted snapshots. Further computational details are given in the Supplementary Methods.

Chemical shift predictions and hydrogen bonding motifs. The predicted chemical shieldings of all snapshots extracted from the MD simulations (168,799,631 total shifts) were obtained using ShiftML version 1.2 (publicly available at <https://shiftml.epfl.ch>)^{29,31}. The conversion from predicted shieldings to isotropic shifts is described in the Supplementary Methods.

H-bonded N-H groups were identified in 11 snapshots from each MD simulation, spaced by 10 ns each. The corresponding bonding motifs were extracted by defining hydrogen bonds as N-H...X (X = O, N) patterns with an N-H-X angle above 130° and H-X bond length shorter than 2.5 Å, typically corresponding to moderate to strong hydrogen bonds in organic solids³⁷. If the first H-bonded neighbour was found to be a water molecule, then secondary water-bound neighbours were searched for using the same criteria to define hydrogen bonds.

In addition, the N-H groups yielding predicted ¹H chemical shifts above 11 ppm were identified within each snapshot of the 4% water MD simulations (2,002 total snapshots), and the corresponding hydrogen bonding patterns were extracted as described above.

Formation energies in the amorphous simulations. The formation energy of the intermolecular complex made of one molecule of AZD5718 and its local environment was computed for each molecule in the same 11 snapshots per simulation as used for identification of all the hydrogen bonding motifs. The environment of a molecule was defined as all molecules having at least one atom within 5 Å from any atom in the probe molecule. The formation energy was computed as the difference between the energy of the total intermolecular complex and the energy of the isolated environment. The obtained formation energy thus contains both the ground-state energy of the isolated probe molecule, which includes its conformational energy, and the interaction energy between the probe molecule and its environment. The single-point energy computations were performed at the DFTB3-D3H5 level of theory using the 3ob-3-1 parameter set and the DFTB + software version 20.1^{38–43}.

Data availability

All data used in this study is freely available in the Materials Cloud repository, <https://doi.org/10.24435/materialscloud-gg-mx>. Source data are provided with this paper.

Received: 4 December 2020; Accepted: 25 March 2021;

Published online: 20 May 2021

References

- Harris, K. D. M. Powder diffraction crystallography of molecular solids. *Top. Curr. Chem.* **315**, 133–177 (2012).
- Hughes, C. E., Boughdiri, I., Bouakkaz, C., Williams, P. A. & Harris, K. D. M. Elucidating the crystal structure of dl-arginine by combined powder X-ray diffraction data analysis and periodic DFT-D calculations. *Cryst. Growth Des.* **18**, 42–46 (2017).
- Baia, M. et al. De novo determination of the crystal structure of a large drug molecule by crystal structure prediction-based powder NMR crystallography. *J. Am. Chem. Soc.* **135**, 17501–17507 (2013).
- Ashbrook, S. E. & McKay, D. Combining solid-state NMR spectroscopy with first-principles calculations—a guide to NMR crystallography. *Chem. Commun.* **52**, 7186–7204 (2016).
- Reif, B., Ashbrook, S. E., Emsley, L. & Hong, M. Solid-state NMR spectroscopy. *Nat. Rev. Methods Primers* **1** (2021).
- Amstad, E. et al. Production of amorphous nanoparticles by supersonic spray-drying with a microfluidic nebulator. *Science* **349**, 956–960 (2015).
- Descamps, M. Amorphous pharmaceutical solids. *Adv. Drug Deliv. Rev.* **100**, 1–2 (2016).
- Nartowski, K. P. et al. 19F NMR spectroscopy as a highly sensitive method for the direct monitoring of confined crystallization within nanoporous materials. *Angew. Chem. Int. Ed.* **55**, 8904–8908 (2016).
- Rams-Baron, M. et al. in *Amorphous drugs: benefits and challenges* 1–7 (Springer International Publishing, 2018).
- Suresh, K. & Matzger, A. J. Enhanced drug delivery by dissolution of amorphous drug encapsulated in a water unstable metal–organic framework (MOF). *Angew. Chem.* **131**, 16946–16950 (2019).
- Wilson, V. R. et al. Amorphous solid dispersions of enzalutamide and novel polysaccharide derivatives: investigation of relationships between polymer structure and performance. *Sci. Rep.-Uk* **10** (2020).
- Tong, P. & Zografi, G. Effects of water vapor absorption on the physical and chemical stability of amorphous sodium indomethacin. *Aaps PharmSciTech* **5**, 9–16 (2004).
- Surana, R., Pyne, A. & Suryanarayanan, R. Effect of aging on the physical properties of amorphous trehalose. *Pharm. Res.* **21**, 867–874 (2004).
- Indulkar, A. S., Lou, X. C., Zhang, G. G. Z. & Taylor, L. S. Insights into the dissolution mechanism of ritonavir-copovidone amorphous solid dispersions: importance of congruent release for enhanced performance. *Mol. Pharm.* **16**, 1327–1339 (2019).
- Brown, S. P. Applications of high-resolution 1H solid-state NMR. *Solid State Nucl. Magn. Reson.* **41**, 1–27 (2012).
- Rossini, A. J. et al. Dynamic nuclear polarization enhanced nmr spectroscopy for pharmaceutical formulations. *J. Am. Chem. Soc.* **136**, 2324–2334 (2014).
- Li, M. et al. Understanding molecular interactions in raxofenamide-povidone amorphous solid dispersions from ultrafast magic angle spinning NMR. *Mol. Pharm.* **17**, 2196–2207 (2020).

18. Baías, M. et al. Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state H-1 NMR spectroscopy. *Phys. Chem. Chem. Phys.* **15**, 8069–8080 (2013).
19. Ruiz-Preciado, M. A. et al. Supramolecular modulation of hybrid perovskite solar cells via bifunctional halogen bonding revealed by two-dimensional F-19 solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **142**, 1645–1654 (2020).
20. Pickard, C. J. & Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B* **63**, 245101 (2001).
21. Charpentier, T. The PAW/GIPAW approach for computing NMR parameters: a new dimension added to NMR study of solids. *Solid State Nucl. Magn. Reson.* **40**, 1–20 (2011).
22. Hartman, J. D. & Beran, G. J. Fragment-based electronic structure approach for computing nuclear magnetic resonance chemical shifts in molecular crystals. *J. Chem. Theory Comput.* **10**, 4862–4872 (2014).
23. Hartman, J. D., Monaco, S., Schatschneider, B. & Beran, G. J. Fragment-based (13)C nuclear magnetic resonance chemical shift predictions in molecular crystals: an alternative to planewave methods. *J. Chem. Phys.* **143**, 102809 (2015).
24. Hartman, J. D., Kudla, R. A., Day, G. M., Mueller, L. J. & Beran, G. J. Benchmark fragment-based (1)H, (13)C, (15)N and (17)O chemical shift predictions in molecular crystals. *Phys. Chem. Chem. Phys.* **18**, 21686–21709 (2016).
25. Hartman, J. D., Day, G. M. & Beran, G. J. O. Enhanced NMR discrimination of pharmaceutically relevant molecular crystal forms through fragment-based ab initio chemical shift predictions. *Cryst. Growth Des.* **16**, 6479–6493 (2016).
26. Xiang, T. X. & Anderson, B. D. Molecular dynamics simulation of amorphous indomethacin. *Mol. Pharm.* **10**, 102–114 (2013).
27. Xiang, T.-X. & Anderson, B. D. Molecular dynamics simulation of amorphous indomethacin-poly(vinylpyrrolidone) glasses: solubility and hydrogen bonding interactions. *J. Pharm. Sci.* **102**, 876–891 (2013).
28. Yuan, X., Xiang, T.-X., Anderson, B. D. & Munson, E. J. Hydrogen bonding interactions in amorphous indomethacin and its amorphous solid dispersions with poly(vinylpyrrolidone) and poly(vinylpyrrolidone-co-vinyl acetate) studied using 13C solid-state NMR. *Mol. Pharm.* **12**, 4518–4528 (2015).
29. Paruzzo, F. M. et al. Chemical shifts in molecular solids by machine learning. *Nat. Commun.* **9**, 1–10 (2018).
30. Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. The Cambridge structural database. *Acta Crystallogr. B* **72**, 171–179 (2016).
31. Engel, E. A. et al. A Bayesian approach to NMR crystal structure determination. *Phys. Chem. Chem. Phys.* **21**, 23385–23400 (2019).
32. Hofstetter, A. & Emsley, L. Positional variance in NMR crystallography. *J. Am. Chem. Soc.* **139**, 2573–2576 (2017).
33. Brown, S. P. & Spiess, H. W. Advanced solid-state NMR methods for the elucidation of structure and dynamics of molecular, macromolecular, and supramolecular systems. *Chem. Rev.* **101**, 4125–4156 (2001).
34. Melnyk, A. et al. Macroscopic structural compositions of π -conjugated polymers: combined insights from solid-state NMR and molecular dynamics simulations. *J. Phys. Chem. Lett.* **8**, 4155–4160 (2017).
35. Schahl, A., Gerber, I. C., Réat, V. & Jolibois, F. Diversity of the hydrogen bond network and its impact on NMR parameters of amylose B polymorph: a study using molecular dynamics and DFT calculations within periodic boundary conditions. *J. Phys. Chem. B* **125**, 158–168 (2020).
36. Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **65**, 1604–1608 (1996).
37. Steiner, T. The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.* **41**, 48–76 (2002).
38. Elstner, M. et al. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B* **58**, 7260–7268 (1998).
39. Aradi, B., Hourahine, B. & Frauenheim, T. DFTB+, a sparse matrix-based implementation of the DFTB method. *J. Phys. Chem. A* **111**, 5678–5684 (2007).
40. Gaus, M., Cui, Q. A. & Elstner, M. DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB). *J. Chem. Theory Comput.* **7**, 931–948 (2011).
41. Gaus, M., Goez, A. & Elstner, M. Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.* **9**, 338–354 (2013).
42. Rezac, J. Empirical self-consistent correction for the description of hydrogen bonds in DFTB3. *J. Chem. Theory Comput.* **13**, 4804–4817 (2017).
43. Hourahine, B. et al. DFTB plus, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **152**, 124101 (2020).

Acknowledgements

We would like to acknowledge Anna Pettersen (AstraZeneca) for single crystal growth and Mark E. Light at the University of Southampton for determining the structure of crystalline **1** Form A using single-crystal X-ray diffraction data. This work was supported by AstraZeneca, Swiss National Science Foundation Grant No. 200020_178860, and by the NCCR MARVEL.

Author contributions

M.B., P.B. and F.P. performed the NMR experiments. S.N.L. performed the crystal structure prediction. M.C. and B.S. computed DFT chemical shifts on candidate crystal structures. M.C. performed the positional uncertainty analysis. E.E. ran the molecular dynamics simulations. A.H. predicted chemical shifts of the MD structures. M.C. computed intermolecular complex formation energies. A.A., M.Q., S.N. prepared and chemically characterized the samples. S.S. and L.E. conceived and supervised the research, and analyzed the results. M.C., M.B., S.S. and L.E. wrote the manuscript with the contribution of all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-23208-7>.

Correspondence and requests for materials should be addressed to S.S. or L.E.

Peer review information *Nature Communications* thanks Eric J. Munso, Joseph Lubach and the other, anonymous, reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021