

# Atomic-level structure determination of amorphous molecular solids by NMR

Received: 8 May 2023

Accepted: 10 August 2023

Published online: 23 August 2023

Check for updates

Manuel Cordova<sup>1,2</sup>, Pinelopi Moutzouri<sup>1</sup>, Sten O. Nilsson Lill<sup>3</sup>, Alexander Cousen<sup>4</sup>, Martin Kearns<sup>5</sup>, Stefan T. Norberg<sup>6</sup>, Anna Svensk Ankarberg<sup>6</sup>, James McCabe<sup>5</sup>, Arthur C. Pinon<sup>7</sup>, Staffan Schantz<sup>6</sup> & Lyndon Emsley<sup>1,2</sup>

Structure determination of amorphous materials remains challenging, owing to the disorder inherent to these materials. Nuclear magnetic resonance (NMR) powder crystallography is a powerful method to determine the structure of molecular solids, but disorder leads to a high degree of overlap between measured signals, and prevents the unambiguous identification of a single modeled periodic structure as representative of the whole material. Here, we determine the atomic-level ensemble structure of the amorphous form of the drug AZD4625 by combining solid-state NMR experiments with molecular dynamics (MD) simulations and machine-learned chemical shifts. By considering the combined shifts of all <sup>1</sup>H and <sup>13</sup>C atomic sites in the molecule, we determine the structure of the amorphous form by identifying an ensemble of local molecular environments that are in agreement with experiment. We then extract and analyze preferred conformations and intermolecular interactions in the amorphous sample in terms of the stabilization of the amorphous form of the drug.

Structure-activity relations drive most areas of modern chemistry. For example, the design of efficient and safe pharmaceutical drugs can be rationalized through the understanding of their atomic-level structure. This can greatly accelerate the search for new compounds with specific properties<sup>1–3</sup>. Tools to determine atomic-level structures have thus become a vital part of modern chemistry research. This is a particular challenge for powdered molecular solids.

In contrast to methods such as powder X-ray diffraction<sup>4–6</sup> or electron diffraction<sup>7–12</sup>, NMR directly probes the local atomic environment, allowing for structural characterization without the need for long-range order<sup>13</sup>. In this direction, solid-state NMR has seen spectacular progress in the last few years<sup>13–16</sup>, and methods have been introduced to solve crystal structures of bulk inorganic<sup>17–20</sup> or

molecular solids<sup>14,15,21–28</sup>. This has resulted in successful structure determination of a variety of powdered materials<sup>13</sup>, including organic solids<sup>14,23–25,29–34</sup>, enzyme active sites<sup>35</sup>, cementitious materials<sup>36–38</sup>, battery materials<sup>39,40</sup>, and hybrid perovskite materials<sup>41</sup>. These structures have been solved by comparing density functional theory (DFT) chemical shifts (or other NMR parameters) computed on model structures (typically generated through crystal structure prediction (CSP) protocols) with experimental values<sup>22–25</sup>.

Despite these remarkable results, complete atomic-level structure determination of amorphous molecular solids remains extremely challenging<sup>42,43</sup>. Nevertheless, amorphous solids are becoming increasingly important. For example, the development of amorphous drug formulations is of current high interest in the pharmaceutical

<sup>1</sup>Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. <sup>2</sup>National Centre for Computational Design and Discovery of Novel Materials MARVEL, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland. <sup>3</sup>Data Science & Modelling, Pharmaceutical Sciences, R&D, AstraZeneca, Gothenburg, Sweden. <sup>4</sup>Early Chemical Development, Pharmaceutical Sciences, R&D, AstraZeneca, Macclesfield, UK. <sup>5</sup>Early Product Development and Manufacturing, Pharmaceutical Sciences, R&D, AstraZeneca, Macclesfield, UK. <sup>6</sup>Oral Product Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, Gothenburg, Sweden. <sup>7</sup>Swedish NMR Center, Department of Chemistry and Molecular Biology, University of Gothenburg, 41390 Gothenburg, Sweden. ✉e-mail: [Staffan.Schantz@astrazeneca.com](mailto:Staffan.Schantz@astrazeneca.com); [lyndon.emsley@epfl.ch](mailto:lyndon.emsley@epfl.ch)

industry, owing to their enhanced solubility and bioavailability with respect to crystalline drugs<sup>44–47</sup>. However, in the absence of methods for atomic-level structure determination, it is not possible to rationalize the factors that lead to the stabilization of amorphous forms, which is a crucial step in developing stable formulations.

The disorder inherent to these compounds leads to the broadening of NMR signals, which leads to significant overlap between the peaks associated with different atomic sites. Consequently, this increases the need for multi-dimensional experiments, which are more difficult to obtain than for crystalline materials due to the lower sensitivity associated with broader lineshapes. The assignment of chemical shifts for amorphous compounds is thus often challenging. Recent advances in dynamic nuclear polarization (DNP)<sup>48–50</sup> have resulted in significant gains of sensitivity in crystalline and amorphous molecular solids, leading to a significant reduction in experimental time required to obtain multi-dimensional NMR spectra of solids.

In addition to these experimental considerations, modeling amorphous structures of materials generally requires the use of molecular dynamics (MD) simulations of large cells typically containing hundreds of molecules. This results in a prohibitive cost for computing chemical shifts using DFT for such large systems. Several approaches have been introduced in order to circumvent this drawback, ranging from using small (hundreds of atoms) amorphous system sizes<sup>36–38,51,52</sup> to isolating local environments to compute chemical shift<sup>35,53,54</sup> to including the effect of long-range interactions by approximate methods<sup>55–59</sup>. While these methods do enable the computation of chemical shifts at the DFT level of theory for amorphous solids, the computational cost remains significant, preventing large-scale chemical shift computations.

Structural disorder has been investigated in proteins by a combination of solid-state NMR, structure generation algorithms, and chemical shift predictions<sup>60–62</sup>. However, such studies have relied on models of chemical shifts in proteins based in part on their primary and/or secondary structure<sup>63–66</sup>. Such models are thus not directly applicable to other molecular solids.

Machine learning (ML) models developed in recent years have proven able to reproduce quantum mechanical properties of materials with similar accuracy as DFT, and at a fraction of the computational cost<sup>67,68</sup>. In particular, ML models of chemical shifts have been introduced and shown to be as accurate as DFT<sup>64,65,69–78</sup>. Recently, we introduced ShiftML, an ML model trained to predict chemical shifts in molecular solids<sup>79</sup>. In its most recent version, the model is trained to reproduce DFT results for solids containing up to 12 elements, and includes distorted geometries, which would be key to describing amorphous systems<sup>80</sup>.

We previously showed how combining MD simulations with large-scale chemical shift predictions obtained using ShiftML allowed the determination of the hydrogen bonding structures in an amorphous drug by comparison with experimentally obtained shifts<sup>42</sup>. However, this approach used a single chemical shift, to focus on the determination of the hydrogen bonding motifs in the structure, as a proof of concept.

Here, we determine the complete ensemble atomic-level structure of the amorphous drug AZD4625<sup>81,82</sup> through the combination of DNP-enhanced solid-state NMR, MD, and machine-learned chemical shifts. To do this we introduce a general approach that integrates multiple chemical shifts and includes the experimental spread of chemical shift distributions in NMR spectra of molecular solids, that we use to select an ensemble of local molecular environments that best match the chemical shift distributions in the measured spectra. This process is applied to over one million molecules from MD simulations, for which we predict chemical shifts. From an analysis of the extracted ensemble of local molecular environments in best agreement with the experiments, we identify key intermolecular interactions and conformations present in the amorphous sample. The local atomic

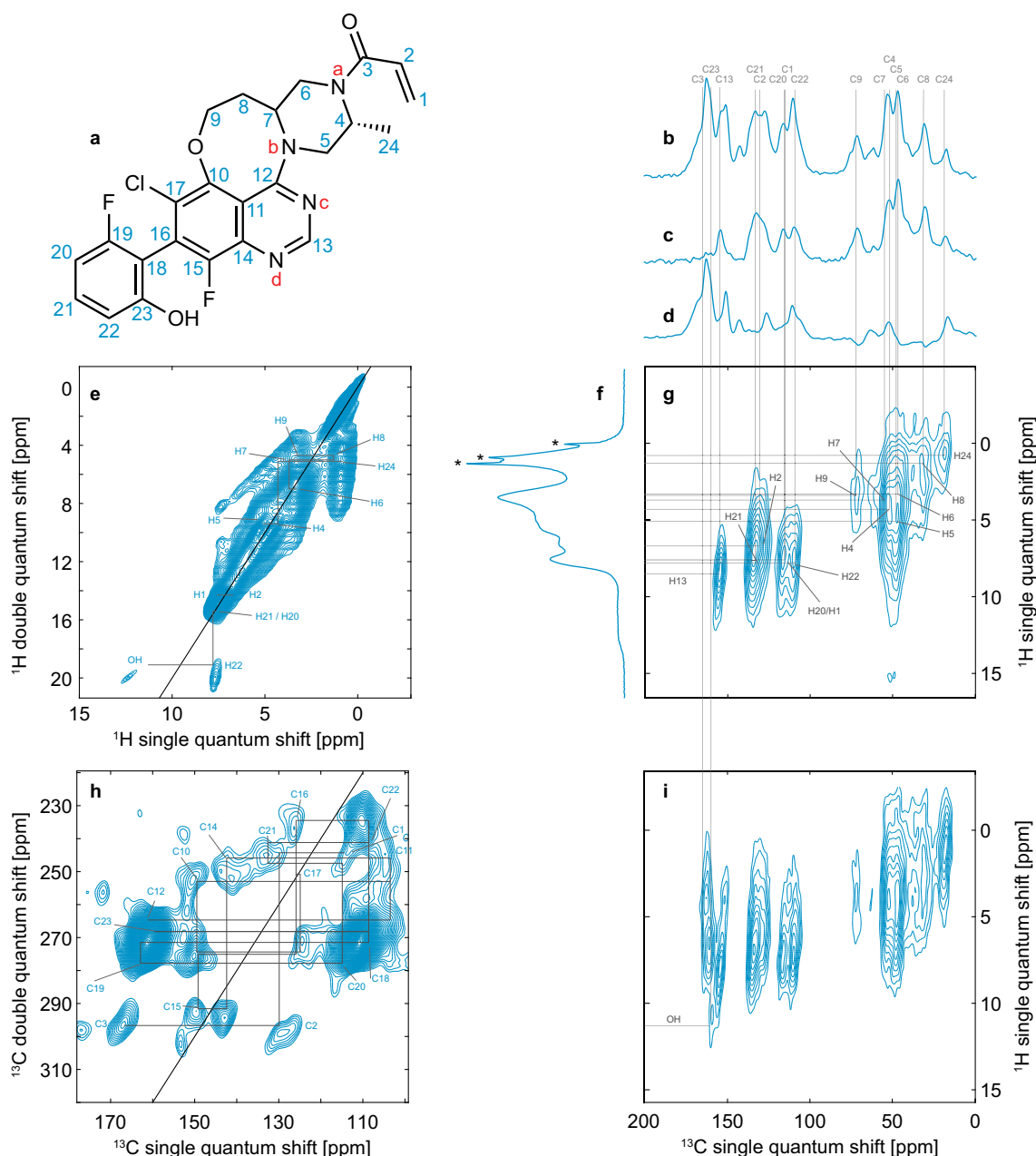
environments determined by NMR were found to accurately reproduce the radial distribution function measured for the sample by powder X-ray diffraction, and to correspond to energetically favorable local structures.

## Results

Figure 1 shows the chemical structure of AZD4625 and the labeling scheme used here, as well as the experimental 1D and 2D NMR spectra obtained for the amorphous form of AZD4625. The spectra display broad linewidths, typical of disordered systems. This highlights the need for multi-dimensional experiments in order to obtain a confident assignment, by spreading the signals over multiple dimensions. With this set of spectra the <sup>1</sup>H and <sup>13</sup>C chemical shifts obtained were assigned as described in the Methods section, leading to the assignments given in Supplementary Table 6. By fitting Gaussian functions to resolved peaks in the 1D <sup>1</sup>H and <sup>13</sup>C MAS spectra, and 2D <sup>1</sup>H-<sup>1</sup>H DQ/SQ spectrum, we obtained linewidths between 2 and 6 ppm for <sup>13</sup>C, 0.6 and 1 ppm for C-H protons, and 1.8 ppm for the OH proton (see Supplementary Table 6 and Supplementary Figs. 2–5). Here, we assume Gaussian shapes for all experimental distributions of chemical shifts. The extracted experimental chemical shift distributions will then serve as the basis to score molecular environments as described in the Methods section. We note that no crystalline form of pure AZD4625 has previously been reported.

To generate a broad ensemble of possible structures, eight MD simulations were carried out with cells containing 128 molecules of AZD4625, randomly initialized in order to model the amorphous system, as described in the Methods section. Chemical shift predictions performed using ShiftML2 were then compared with the experimental values obtained for <sup>1</sup>H and <sup>13</sup>C (excluding the protons and carbon labeled 1 in Fig. 1a due to the ambiguity in their assignment). A total of 1,025,280 molecular environments, each comprising a central molecule and all molecules that have at least one atom within 7 Å from any atom of the central molecule (see Methods section), were extracted from the MD snapshots. For each atomic site in the central molecule of a molecular environment, we compute the probability that the predicted shift is drawn from the corresponding experimental chemical shift distribution. The probabilities across all atomic sites are then combined into a global probability that the local molecular environment matches the NMR experiments. More details are given in the Methods section. Fig. 2a shows the root-mean-square error (RMSE) between <sup>1</sup>H and <sup>13</sup>C chemical shifts computed for all AZD4625 molecules in each of the 8010 snapshots taken from the MD trajectories, as well as the calculated probability that the local molecular environment of each molecule is consistent with the NMR experiments. This includes the computation of chemical shifts for over a million molecules. As expected, higher probability is correlated with lower <sup>1</sup>H and <sup>13</sup>C shift RMSE, but it is very important to note that the RMSEs only consider the difference between the center of the experimental distributions of shifts, and the corresponding chemical shift prediction for each atomic site, while the probability calculated using Eqs. 1–4 also take into account the width of the experimental distributions as well as the prediction uncertainty, providing an improved picture of the compatibility of a given local molecular environment with the experiments. The histogram of all probabilities of local molecular environments ( $p_j$ ) to match the experiments is shown in Fig. 2b. Here, we selected the 1% of local molecular environments in best agreement with the experiment to construct the NMR ensemble, which corresponds to the probabilities above 33%, as indicated by the dashed vertical line in Fig. 2b.

Here, we independently select molecular environments compatible with the NMR experiments. The generation of environments through the MD simulations is inherently biased by the force field used and the starting configurations. The selection of the subset that best matches the experimental data does not aim here to reproduce the



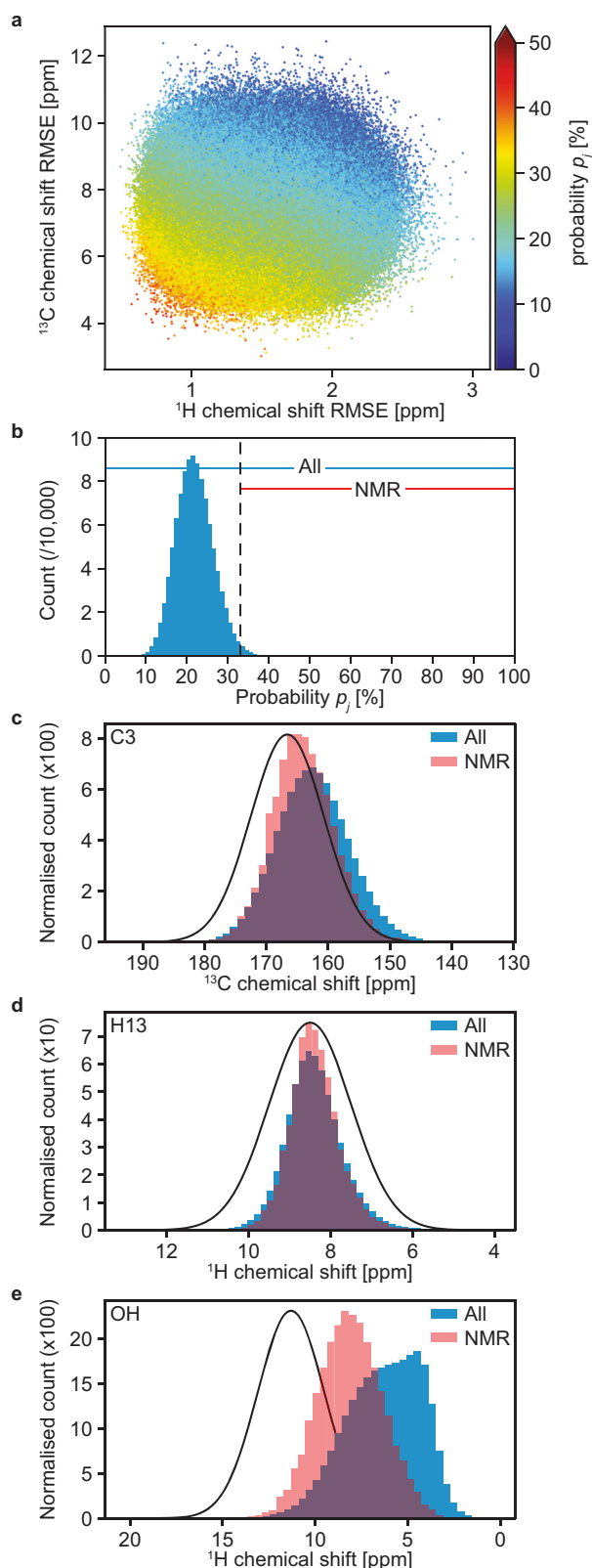
**Fig. 1 | NMR spectra of the amorphous form of AZD4625 used for chemical shift assignment.** **a** Chemical structure of AZD4625 and carbon (blue numbers) and nitrogen (red letters) labeling schemes used here. **1D** **(b–d)** DNP-enhanced  $^{13}\text{C}$  CPMAS spectra without **(b)**, **(c)** and with **(d)** CPPI spectral editing. **f**  $^1\text{H}$  100 kHz MAS spectrum. **2D** **(e)**  $^1\text{H}$ - $^{13}\text{C}$  DQ/SQ, **(g, i)** DNP-enhanced  $^1\text{H}$ - $^{13}\text{C}$  DUMBO-HETCOR

and **(h)** DNP-enhanced  $^{13}\text{C}$ - $^{13}\text{C}$  INADEQUATE spectra of amorphous AZD4625. In **d**  $-\text{CH}_2$  groups appear negative,  $-\text{CH}$  groups disappear, and  $-\text{C}$  and  $-\text{CH}_3$  groups retain a positive intensity. The gray lines indicate correlated peaks or  $^{13}\text{C}$  chemical shifts of protonated carbon species. The stars in **f** indicate artifacts due to mobile impurities in the rotor.

exact experimental ensemble of molecular environments in the sample (as is done, e.g., in NMR studies of intrinsically disordered proteins<sup>83–85</sup>), but here it provides an additional bias in order to identify systematic structural differences from the ensemble generated by MD, as seen below.

Figure 2c–e shows the histograms of chemical shifts computed for carbon labeled 3, proton labeled 13, and of the OH proton for all AZD4625 molecules in the MD trajectories as compared to those from the NMR ensemble. These examples are taken to illustrate the typical changes of chemical shift distributions seen upon the selection of local atomic environments. The distributions for all other protons and carbons considered are given in Supplementary Figs. 7–11. The

distribution of predicted shifts for carbon labeled 3 (Fig. 2c) was found to be significantly closer to the experimental distribution of shifts upon selection of local molecular environments, suggesting that this chemical shift does discriminate between the structures. In contrast, for example, the distribution of predicted shifts for the proton labeled 13 (Fig. 2d), which already displays a large overlap with the corresponding experimental distribution of shifts, does not display a significant change upon the selection of local molecular environments. Then we note that the distribution of predicted chemical shifts for the OH proton (Fig. 2e) displays a large difference after the selection of local molecular environments, again suggesting that this shift is a powerful discriminator. However, even after the selection of the best



**Fig. 2 | Ensemble structure determination.** **a** A comparison of  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift RMSEs for each molecule in the MD snapshots, colored according to its probability to be simultaneously compatible with the experimental shift distributions for all assigned atoms (as described by Eqs. 1–4 in the Methods section). **b** Histogram of the probabilities of all molecules in the MD snapshots to be compatible with the experimental shift distributions. The dashed line indicates the probability threshold used to select local molecular environments. The ranges of probabilities included in the whole and the NMR ensembles are indicated above the histogram. Examples of the predicted chemical shift distributions for the (c) carbon labeled 3, (d) proton labeled 13, and (e) OH proton in all molecular environments (blue) in the MD snapshots and in the NMR ensemble (red), compared to the corresponding experimentally measured distributions (black). Equivalent figures for all the other assigned atoms are given in Supplementary Figs. 7–11.

Figure 3 shows the analysis of structural properties in the set of best-match molecular environments, compared to all molecular environments present in all MD snapshots. As seen in Fig. 3a, the selection of local molecular environments compatible with the NMR experiments promotes hydrogen bonds, in particular with the oxygen labeled 3 and the nitrogen labeled c. Accordingly, the proportion of OH protons not forming hydrogen bonds is significantly reduced in the set of selected local molecular environments. Hydrogen bonding to nitrogen was found to generally lead to further deshielding of the OH proton compared to hydrogen bond to oxygen, as seen in Supplementary Fig. 12.

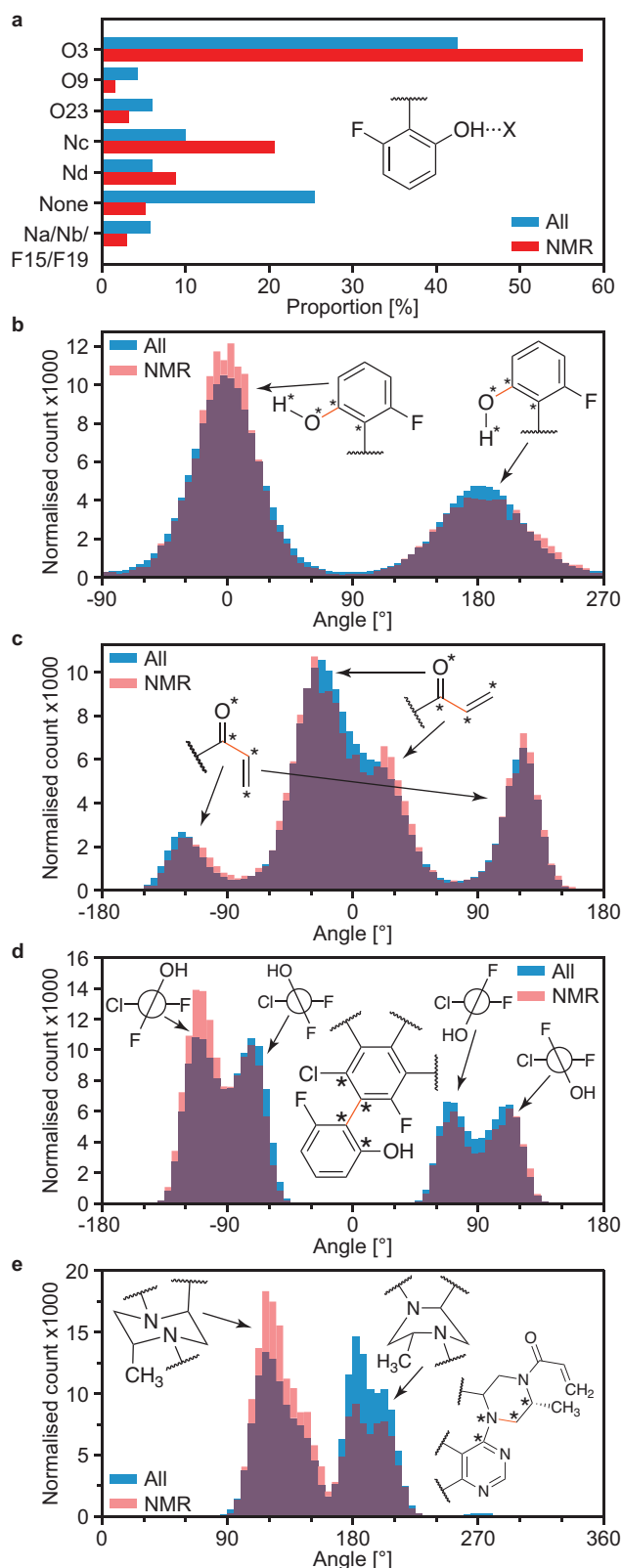
Preferred conformations of AZD4625 can be extracted from the NMR ensemble. Figure 3b shows that the position of the OH proton is generally preferred to be pointing away from the body of the molecule, and that this trend is slightly reinforced in the NMR ensemble. Similarly, the Z conformation of the enone group is found to be preferred, and that preference is retained in the NMR ensemble (Fig. 3c). The conformation yielding dihedral angles between the aromatic planes from  $-120$  to  $-90^\circ$  were found to be promoted in the NMR ensemble (Fig. 3d). We note that for this case, five of the eight MD simulations carried out started with a dihedral angle around  $-90^\circ$  and three of them started with an angle around  $90^\circ$ , which explains the difference in the height of the distributions for positive and negative values in all molecules from the MD snapshots (more details are given in SI). The chair conformation of the aromatic 6-membered ring was also found to be promoted by the NMR selection of local molecular environments compared to the boat conformation that was also observed in the MD simulations (Fig. 3e).

It is interesting to compare the total radial distribution function  $G(r)$  and differential correlation function  $D(r)$  obtained from the ensembles before and after the selection of local molecular environments with the functions obtained experimentally by powder X-Ray diffraction (Fig. 4). The MD trajectories were found to accurately reproduce the experimental data, with the largest differences found in the two peaks at 1.4 and 2.4 Å. This can be attributed to differences in bond lengths between the MD simulations and the sample. Importantly, the features at distances above 3 Å are correctly captured by the MD simulation. The selection of local molecular environments was not found to significantly change the similarity between the simulated and experimental  $G(r)$  or  $D(r)$ . This result highlights that the scattering data is unable to sensitively discriminate between ensembles of local molecular environments in the samples studied here.

Figure 5 shows the predicted formation energies of molecules of AZD4625 with their local environment, including the formation energy of the central molecule (as described in the Methods section). This is a measure of the stabilization of the molecules by their environment. The local environments in the NMR ensemble were found to result on average in stabilization of the central molecule as compared to random local molecular environments extracted from the MD simulations, by  $8.7 \pm 0.7$  kJ/mol on average (Fig. 5a). This result suggests that the selection of molecular environments,

match structures, the overlap with the predicted distribution is not perfect. We attribute this to the significant proportion of OH protons weakly bonded to hydrogen bond acceptors in the MD trajectories (see Supplementary Fig. 12). This effect may also be due to bias in the shift predictions. We also note that importantly the best match selection does not critically depend on any single shift, but is the result of the joint match to all the shifts in the molecule.





**Fig. 3 | Structural properties of the amorphous form of AZD4625.** **a** Proportions of different hydrogen bond acceptors bonded to the OH group of AZD4625 in all local molecular environments (blue) and in the NMR ensemble (red). Histogram of dihedral angles for the **b** OH group, **c** enone, **d** aromatic planes, and **e** aliphatic ring in all molecules (blue) and in the NMR ensemble (red). In **b–e** the rotatable bond associated with the dihedral angle is drawn in orange. Stars indicate the atoms used for the computation of the dihedral angle.

based purely on NMR chemical shifts, also led to the selection of energetically favorable local molecular environments. Figure 5b shows that hydrogen bonding of the OH proton of a central molecule to either oxygen labeled 3 or nitrogen labeled d leads to enhanced stabilization of the central molecule by its whole environment. This also corroborates the increase in hydrogen bonds formed with these two atoms in the NMR ensemble of molecular environments discussed above (Fig. 3a).

A set of 20 randomly selected central molecules from the NMR ensemble is shown in Fig. 6a. This highlights the structural flexibility of AZD4625 in the amorphous state. Fig. 6b shows three-dimensional atomic density maps around the OH proton in the NMR (left panel) and the random (middle panel) local molecular environments, as well as the difference between the two atomic density maps (right panel). As expected from Figs. 3a and 5b, hydrogen bonding towards oxygen and nitrogen atoms is promoted by the selection of local molecular environments. This is highlighted by the contours representing nitrogen and oxygen atomic densities in the rightmost panel in Fig. 6b. This suggests that these interactions are critical to stabilizing the structure of amorphous AZD4625. Figure 6c shows similar atomic density maps, aligned around the methyl group of AZD4625. The difference between atomic density maps highlights the preferred conformation of the 6- and 8-membered aliphatic rings.

## Discussion

We have determined the ensemble atomic-level structure of the amorphous form of AZD4625 by combining solid-state NMR experiments with MD simulations and prediction of chemical shifts for over one million AZD4625 molecules in the MD trajectories. Importantly, no crystalline structure of the pure compound has previously been reported.

Local molecular environments compatible with the NMR spectra measured were selected through a general approach that integrates multiple chemical shifts, and includes the spread of chemical shift distributions in the experimental spectra as well as the uncertainty of the chemical shift predictions. We expect that the method presented here can be straightforwardly applied to determine the structure of any molecular solid.

The local atomic environments determined by NMR were found to accurately reproduce the radial distribution function measured for the sample by powder X-Ray diffraction. The NMR ensemble was also found to lead to an overall stabilization of the selected molecules by their environment.

The ensemble of selected local molecular environments highlights key structural properties in the amorphous sample that play a critical role in the structure and stabilization of the material in its amorphous form.

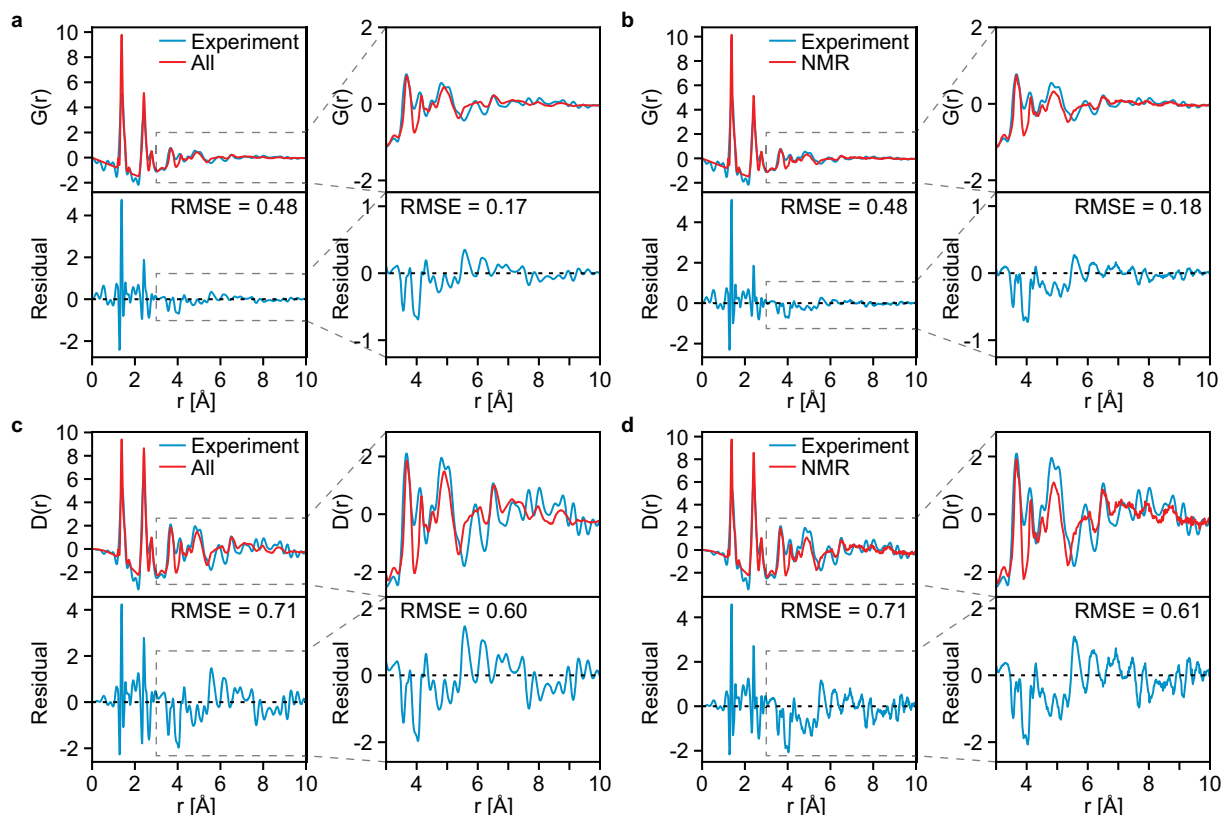
## Methods

### Synthesis

The synthesis of AZD4625 is described in ref. 81. The amorphous AZD4625 solid was precipitated from 2-methyltetrahydrofuran (2-MeTHF) and n-heptane. Crude API was initially dissolved in 2-MeTHF, the solution of which was charged directly to n-heptane at 18 °C. The precipitate was isolated under vacuum and dried from 25–70 °C.

### X-ray diffraction experiments

Synchrotron X-ray PDF data were collected on the I15-1 beamline at Diamond Light Source, UK. Powdered samples were contained within a 1 mm inner diameter polyimide capillary with a 0.025 mm wall thickness and spun perpendicular to the beam during data collection. An empty capillary was also collected for background subtraction. Scattering data were collected at an incident X-ray energy of 76.69 keV with one Perkin Elmer XRD4343CT area detector placed close to the sample (~200 mm) for PDF data and a second Perkin Elmer XRD1611CP3 area



**Fig. 4 | Radial distribution functions.** The total radial distribution function  $G(r)$  (a, b) and the differential correlation function  $D(r)$  (c, d) measured from powder X-ray diffraction (see methods) (blue) and simulated (red) using (a, c) all molecules and (b, d) the best match ensemble by NMR. The lower panels show the residual

between the experiment and simulations in each case, along with the RMSE obtained. The plots on the right of each panel show the range between 3 and 10 Å, and the RMSE in the corresponding range.

detector placed further from the sample (-850 mm) for higher resolution Bragg data; the precise detector geometries were calculated using DAWN<sup>86</sup> from data collected on a crystalline standard (NIST SRM640c). Total data collection times were 30 min for the PDF data and 2 min for Bragg data. 2D scattering data were corrected for polarization, solid angle, and detector thickness prior to integration to 1D using DAWN<sup>86</sup>. The GudrunX program was then used to perform container background, multiple scattering, Compton scattering, and absorption corrections on data in the range  $0.3 \leq Q \leq 26 \text{ \AA}^{-1}$ , prior to Fourier transform to produce the PDF<sup>87</sup>.

### NMR experiments

Experiments were carried out using either room temperature ultra-fast MAS rate techniques that enhance  $^1\text{H}$  spectral resolution or DNP approaches that enhance the sensitivity of NMR signals. DNP is performed at temperatures of -100 K and relies on the transfer of high electron spin polarization, typically from exogenously added solutions of organic radicals, to nuclei of interest upon microwave irradiation<sup>48,49,88,89</sup>.

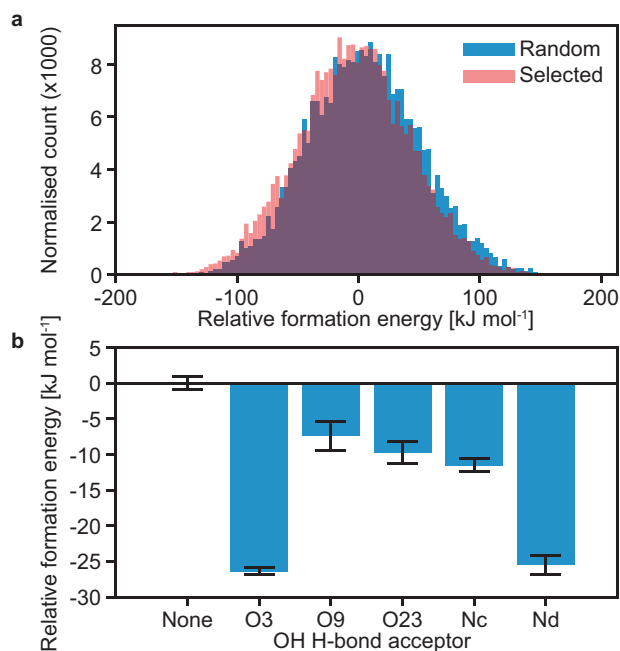
The DNP-enhanced NMR experiments were carried out on commercial Bruker Avance Neo NMR spectrometers at a nominal field strength of 9.40 T equipped with either a 264 GHz klystron or a 263 GHz gyrotron microwave source and a 3.2 mm LTMAS DNP probe in a  $^1\text{H}/^{13}\text{C}/^{15}\text{N}$  configuration which was cooled to about 100 K before sample insertion. The DNP sample was packed into a 3.2 mm sapphire rotor, plugged with a Teflon insert, and topped with a zirconia drive cap. Prior to packing, the powder sample of the amorphous form of AZD4625 was ground by hand in a pestle and mortar and then impregnated<sup>48,49,88,89</sup> with a 20 mM solution of the AMUPol biradical<sup>90</sup> dissolved in a mixture of  $\text{H}_2\text{O}:\text{D}_2\text{O}:\text{C}^{12}\text{-glycerol}$  (10:30:60 v/v). A DNP

enhancement of a factor 6–8 was achieved, measured as the ratio of the ( $^1\text{H}$ ) $^{13}\text{C}$  cross-polarization (CP) signal intensity between spectra acquired with and without microwaves. While this is a modest enhancement, it was sufficient to enable the acquisition of the natural abundance  $^{13}\text{C}$ - $^{13}\text{C}$  INADEQUATE experiments described below. DNP spectra were acquired at MAS rates of 8 or 10 kHz.

The room temperature NMR experiments were performed on a dry sample of the powder at a MAS rate of 100 kHz, using a Bruker 0.7 mm room temperature HCN CPMAS probe at a magnetic field of 21.1 T. A States-TPPI acquisition scheme was used to obtain phase-sensitive two-dimensional spectra. The  $^1\text{H}$  and  $^{13}\text{C}$  chemical shifts were referenced to literature values. More experimental details and a link to the raw NMR data can be found in the SI.

### Chemical shift assignment

The  $^1\text{H}$  and  $^{13}\text{C}$  resonances of the amorphous form of AZD4625 (Fig. 1a) were assigned using one-dimensional  $^1\text{H}$  and  $^{13}\text{C}$  MAS NMR experiments,  $^{13}\text{C}$  CPPI spectral editing<sup>91</sup>, (Fig. 1b–d, f), in combination with two-dimensional  $^1\text{H}$ - $^1\text{H}$ ,  $^{13}\text{C}$ - $^{13}\text{C}$ , and  $^1\text{H}$ - $^{13}\text{C}$  correlation spectra. The  $^1\text{H}$ - $^1\text{H}$  DQ/SQ (Fig. 1e) spectrum provides through-space dipolar correlations between protons, the natural abundance DNP-enhanced refocused  $^{13}\text{C}$ - $^{13}\text{C}$  INADEQUATE<sup>49</sup> (Fig. 1h) provides the covalent connectivities between carbon atoms, and the short- and long-range  $^1\text{H}$ - $^{13}\text{C}$  DNP-Enhanced DUMBO-HETCOR experiments (Fig. 1g, i), provide  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear shift correlations. A DNP-enhanced natural abundance  $^{13}\text{C}$ - $^{13}\text{C}$  INADEQUATE spectrum recorded for a crystalline form was also used to guide the assignment (Supplementary Fig. 1). The chemical shift assignments obtained from an analysis of these spectra for the  $^1\text{H}$  and  $^{13}\text{C}$  nuclei are given in Supplementary Table 6. The chemical shift of C1 was not taken into



**Fig. 5 | Formation energies.** **a** Relative formation energies of intermolecular complexes of 8000 randomly selected molecules (blue) and the molecules from the NMR ensemble (red). The zero is set to be the mean formation energy of all intermolecular complexes. **b** Relative formation energies of the local molecular environments in the NMR ensemble for different hydrogen bond acceptors bonded to the OH proton. The zero is set to be the mean formation energy of intermolecular complexes where no hydrogen bonding acceptor is bonded to the OH proton of the central molecule. Formation energies were computed as the difference in energy between a molecular environment (all molecules with at least one atom within 7 Å from any atom of the central molecule) with and without the central molecule, and thus contain both intermolecular interactions and the conformational energy of the central molecule. The error bars shown are the standard error on the mean.

consideration in the subsequent analysis due to high uncertainty in the assignment.

### MD simulation of AZD4625

The amorphous structure of AZD4625 was modeled by carrying out MD simulations with the OPLS4 force-field<sup>92</sup> in Desmond<sup>93,94</sup> on periodic amorphous cells containing 128 molecules. Eight different amorphous cell simulations were generated and evaluated using Materials Studio<sup>95</sup>. After equilibration for 1 ns using the canonical NVT ensemble first at 100 K and then at 298 K followed by 22 ns using the isothermal-isobaric ensemble (NPT) at 298 K and 1 bar, production simulations were carried out for 500 ns using the NPT ensemble at 298 K and 1 bar. Snapshots of each MD simulation were extracted every 100 ps and input directly to ShiftML2<sup>80</sup> for <sup>1</sup>H and <sup>13</sup>C chemical shift predictions. The chemical shielding values were converted to chemical shifts using offsets of 30.78 and 170.04 ppm for <sup>1</sup>H and <sup>13</sup>C, respectively. The input files of the MD simulation, extracted MD snapshots, and predicted shifts are given with the raw data. Further information about the MD simulations is given in SI.

### Selection of local molecular environments

Local molecular environments, comprising a central molecule and all other molecules having at least one atom within 7 Å from any atomic site in the central molecule, were extracted from the MD snapshots (1,025,280 environments in total) and selected based on the probability of the molecule at the center of each environment to match the experimental distributions of chemical shifts. Considering one atomic

site  $a_i$  in AZD4625, we describe the associated distribution of experimental chemical shifts as a Gaussian function centered on the chemical shift experimentally measured,  $\delta_{\text{exp},a_i}$ , and with a width given by the linewidth of the peaks observed in the spectra,  $\sigma_{\text{exp},a_i}$ . Based on the measurement of the linewidths in the resolved peaks in the spectra of Fig. 1, here we obtained widths between 2 and 6 ppm for the <sup>13</sup>C resonances, and 0.6 and 1 ppm for the <sup>1</sup>H resonances, except for the OH proton for which we obtained a width of 1.8 ppm. The centers and widths of the experimental chemical shift distributions are given in Supplementary Table 6 and Supplementary Figs. 2–5.

The chemical shift  $\delta_{\text{pred},a_i^{(j)}}$  and uncertainty  $\sigma_{\text{pred},a_i^{(j)}}$  predicted using ShiftML2 for that atomic site  $a_i^{(j)}$  in a molecule  $j$  within a given MD snapshot can similarly be described as a Gaussian function centered on the shift prediction and with a width given by the prediction uncertainty. We then define the probability that the computed shift is within the experimental distribution of chemical shift with the two-tailed  $p$  value resulting from the  $Z$  score computed between the two Gaussians:

$$Z_{a_i^{(j)}} = \frac{|\delta_{\text{exp},a_i} - \delta_{\text{pred},a_i^{(j)}}|}{\sqrt{\sigma_{\text{exp},a_i}^2 + \sigma_{\text{pred},a_i^{(j)}}^2}} \quad (1)$$

The  $p$  value  $p_{\text{val}}(Z_{a_i^{(j)}})$  thus corresponds to the probability that the computed shift is drawn from the experimental distribution of chemical shift for that atomic site:

$$p_{\text{val}}(Z_{a_i^{(j)}}) = \sqrt{\frac{2}{\pi}} \cdot \int_{Z_{a_i^{(j)}}}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx. \quad (2)$$

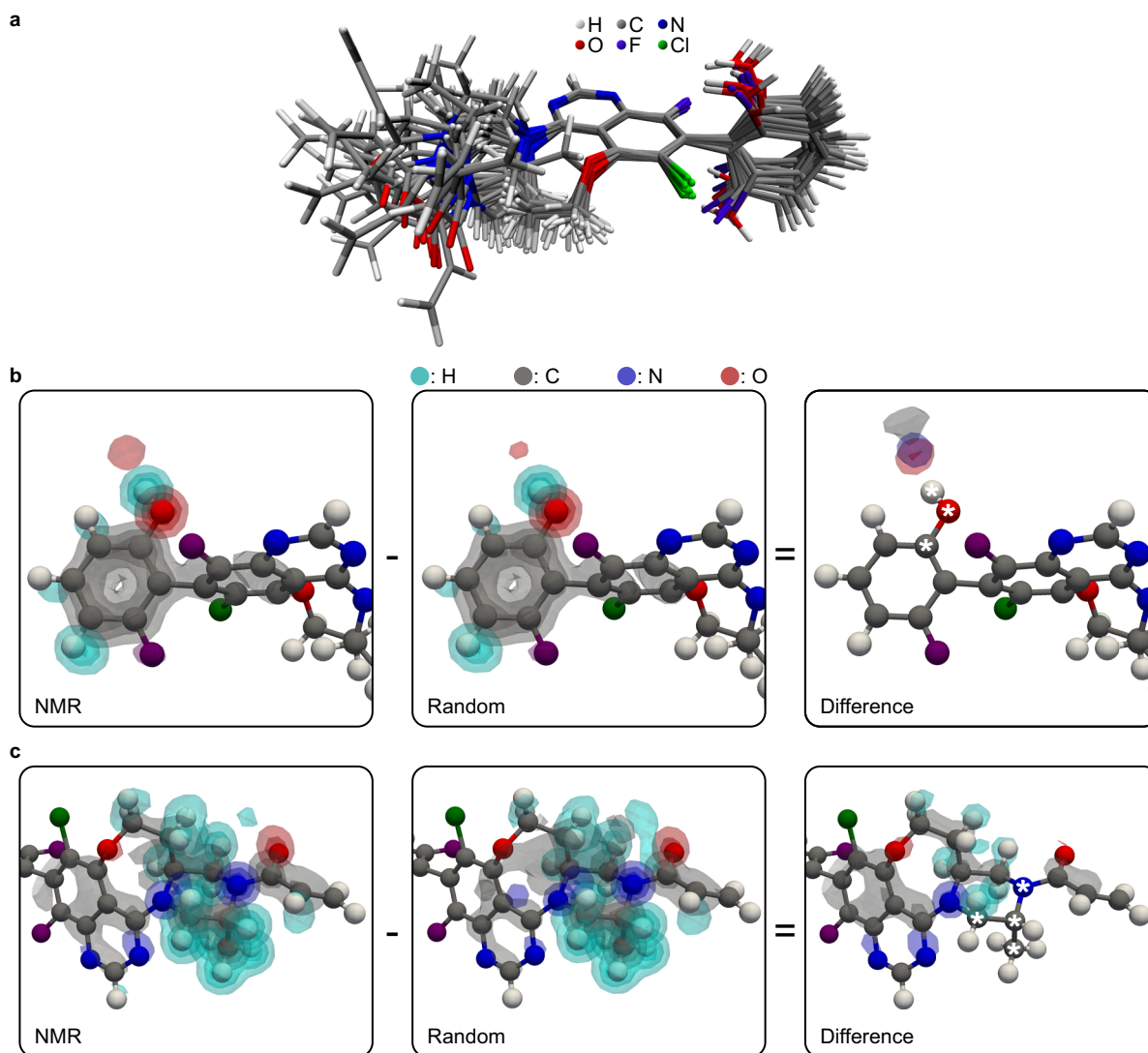
We note that the  $p$  value corresponds to the null hypothesis, which is here that the shift is drawn from the experimental distribution. A large  $p$  value thus indicates a better correspondence between the predicted shift and experimental distribution. To obtain the probability that the computed shift corresponds to the experimental distribution of shifts, we divide the  $p$  value obtained by the prediction uncertainty divided by the first quartile of all predicted uncertainties obtained for that atomic site in all molecules of all MD snapshots,  $\sigma_{\text{pred},a_i}^0$ , capped to a minimum value of 1. This step was done in order to prevent chemical shifts predicted with very high uncertainty, thus where the shift prediction is unreliable, from being artificially associated with a high probability of corresponding to the experimental distribution.

$$p_{a_i^{(j)}} = \frac{p_{\text{val}}(Z_{a_i^{(j)}})}{\max\left(1, \frac{\sigma_{\text{pred},a_i^{(j)}}}{\sigma_{\text{pred},a_i}^0}\right)}. \quad (3)$$

The probability  $p_j$  that a given molecular environment  $j$  within an MD snapshot corresponds to the experimental spectrum was then evaluated as the geometric mean of the probabilities obtained using Eq. 3 for all protons and carbons in the molecule (except, here, for the protons and carbon labeled 1 in Fig. 1a, due to the high uncertainty in the assignment of that carbon). This probability was computed for all local environments in all MD snapshots:

$$p_j = \left(\prod_i^n p_{a_i^{(j)}}\right)^{\frac{1}{n}} \quad (4)$$

The selection of the ensemble of local molecular environments most compatible with the experimental spectra, that we refer to as the NMR ensemble, was then performed by selecting all environments having an overall probability  $p_j$  above 0.33, corresponding to



**Fig. 6 | Structures representative of the molecular conformations present in the amorphous form of AZD4625.** **a** Superposition of 20 molecules of AZD4625 randomly selected from the NMR ensemble. Three-dimensional atomic density maps in NMR-selected and random molecular environments aligned around **(b)** the OH and **(c)** the methyl groups. The difference between the 3D maps for the selected and random molecular environments are shown on the right panels, where the atoms aligned are indicated by asterisks in the difference maps. 3D contours are

drawn at levels of 0.2, 0.4, 0.6, and 0.8 for the atomic density maps and 0.05, 0.1, 0.15, and 0.2 for the difference maps. The conformation of the molecule displayed along with the atomic density map was chosen such that the various dihedral angles best correspond to the maxima of the distributions for selected local molecular environments in Fig. 3b–e. The three-dimensional contours in the rightmost panel in **b** and **c** highlight the overall structural features promoted by the NMR-based selection.

about 1% of all local molecular environments present in the MD snapshots (10,107 environments). We note that the cutoff value of 0.33 was chosen as a balance between the maximization of the overlap and minimization of the Jensen-Shannon divergence<sup>96</sup> with the experimental shift distributions, and the selection of large enough ensemble to describe the amorphous compound (see Supplementary Fig. 6).

In addition, 1000 local molecular environments were randomly selected from each MD simulation to construct a random ensemble for comparison with the experimentally determined ensemble.

#### Computation of formation energies of local molecular environments

The formation energy of local molecular environments was computed as the energy difference between the environments (all molecules with at least one atom within 7 Å from any atom of the central molecule) with and without the central molecule. This energy thus includes both the intermolecular interactions and conformational energy of the

central molecule. The energies were computed using the DFTB-D3H5 semiempirical level of theory using the 3ob-3-1 parameter set and the DFTB+ software version 22.2<sup>97–103</sup>. The computed energies are given with the raw data.

#### Identification of hydrogen bonds in local molecular environments

Hydrogen bonds involving the OH proton of the central molecule in each local molecular environment were identified by defining hydrogen bonds as O–H...X motifs (X = O, N) with an O–H...X angle above 130° and H–X distance shorter than 2.5 Å.

#### Three-dimensional atomic density maps

The three-dimensional atomic density maps were constructed by aligning the selected and random ensembles of local molecular environments on given atoms in the central molecule. This was done by minimizing the root-mean-square displacement between the positions of the atoms used for the alignment in the central molecule of the



different molecular environments. Three-dimensional atomic density maps were then generated by summing three-dimensional Gaussian functions with a width  $\sigma = 0.5 \text{ \AA}$  placed at the atomic positions  $r_{a_i}$  of the aligned local environments, divided by the number of environments aligned.

$$G(\vec{r}) = \frac{1}{N_{env}} \sum_i^{N_{env}} \sum_{a_i \in i} \exp\left(-\frac{|\vec{r} - \vec{r}_{a_i}|^2}{2\sigma^2}\right) \quad (5)$$

Individual atomic density maps were constructed for each element present in the set of aligned environments. The Gaussian functions were not normalized, and this leads to a value of 1 at a given position if an atom of a given element is found at that position in all environments. Each atomic density map was evaluated on a  $31 \times 31 \times 31$  cubic grid centered at the aligned atomic sites and with  $12 \text{ \AA}$  sides. This corresponds to a spatial sampling of  $0.4 \text{ \AA}$ .

## Data availability

The NMR raw data are available from the Materialscloud repository <https://doi.org/10.24435/materialscloud:gk-51> in JCAMP-DX version 6.0 standard format and original TopSpin format, as well as the input files for the MD simulations, the MD snapshots extracted, formation energies of intermolecular complexes, and all scripts used to perform the data analysis. All data and scripts are available under the license CC-BY-4.0 (Creative Commons Attribution-ShareAlike 4.0 International).

## References

- King, R. D., Muggleton, S., Lewis, R. A. & Sternberg, M. J. Drug design by machine learning: the use of inductive logic programming to model the structure-activity relationships of trimethoprim analogues binding to dihydrofolate reductase. *Proc. Natl Acad. Sci.* **89**, 11322–11326 (1992).
- McTigue, M. et al. Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors. *Proc. Natl Acad. Sci.* **109**, 18281–18289 (2012).
- Daina, A., Michielin, O. & Zoete, V. iLOGP: a simple, robust, and efficient description of n-octanol/water partition coefficient for drug design using the GB/SA approach. *J. Chem. Inf. Model.* **54**, 3284–3301 (2014).
- Rietveld, H. M. A profile refinement method for nuclear and magnetic structures. *J. Appl. Crystallogr.* **2**, 65 (1969).
- 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).
- Gruene, T. et al. Rapid structure determination of microcrystalline molecular compounds using electron diffraction. *Angew. Chem. Int Ed.* **57**, 16313–16317 (2018).
- Gemmi, M. et al. 3D electron diffraction: the nanocrystallography revolution. *ACS Cent. Sci.* **5**, 1315–1329 (2019).
- Gruene, T., Holstein, J. J., Clever, G. H. & Keppler, B. Establishing electron diffraction in chemical crystallography. *Nat. Rev. Chem.* **5**, 660–668 (2021).
- Huang, Z. H., Grape, E. S., Li, J., Inge, A. K. & Zou, X. D. 3D electron diffraction as an important technique for structure elucidation of metal-organic frameworks and covalent organic frameworks. *Coord. Chem. Rev.* **427**, 213583 (2021).
- Jones, C. G. et al. The CryoEM method MicroED as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* **4**, 1587–1592 (2018).
- Nannenga, B. L. & Gonen, T. The cryo-EM method microcrystal electron diffraction (MicroED). *Nat. Methods* **16**, 369–379 (2019).
- Reif, B., Ashbrook, S. E., Emsley, L. & Hong, M. Solid-state NMR spectroscopy. *Nat. Rev. Methods Primers* **1**, 2 (2021).
- Hodgkinson, P. NMR crystallography of molecular organics. *Prog. Nucl. Magn. Reson. Spectrosc.* **118–119**, 10–53 (2020).
- Southern, S. A. & Bryce, D. L. In: Annual reports on NMR spectroscopy, Vol. 102 *Annu. Rep. NMR Spectroscopy* (ed. G.A. Webb) 1–80 (2021).
- Kubicki, D. J., Stranks, S. D., Grey, C. P. & Emsley, L. NMR spectroscopy probes microstructure, dynamics and doping of metal halide perovskites. *Nat. Rev. Chem.* **5**, 624–645 (2021).
- Brouwer, D. H. et al. A general protocol for determining the structures of molecularly ordered but noncrystalline silicate frameworks. *J. Am. Chem. Soc.* **135**, 5641–5655 (2013).
- Brouwer, D. H., Darton, R. J., Morris, R. E. & Levitt, M. H. A solid-state NMR method for solution of zeolite crystal structures. *J. Am. Chem. Soc.* **127**, 10365–10370 (2005).
- Loiseau, T. et al. MIL-96, a porous aluminum trimesate 3D structure constructed from a hexagonal network of 18-membered rings and  $\mu(3)$ -oxo-centered trinuclear units. *J. Am. Chem. Soc.* **128**, 10223–10230 (2006).
- 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).
- 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**, 4125–4155 (2001).
- Elena, B. & Emsley, L. Powder crystallography by proton solid-state NMR spectroscopy. *J. Am. Chem. Soc.* **127**, 9140–9146 (2005).
- Salager, E. et al. Powder crystallography by combined crystal structure prediction and high-resolution H-1 solid-state NMR Spectroscopy. *J. Am. Chem. Soc.* **132**, 2564 (2010).
- Baias, 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).
- Baias, 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).
- Brus, J. et al. Predicting the crystal structure of decitabine by powder NMR crystallography: influence of long-range molecular packing symmetry on NMR parameters. *Cryst. Growth Des.* **16**, 7102–7111 (2016).
- Balodis, M., Cordova, M., Hofstetter, A., Day, G. M. & Emsley, L. De novo crystal structure determination from machine learned chemical shifts. *J. Am. Chem. Soc.* **144**, 7215–7223 (2022).
- Hofstetter, A. et al. Rapid structure determination of molecular solids using chemical shifts directed by unambiguous prior constraints. *J. Am. Chem. Soc.* **141**, 16624–16634 (2019).
- Czernek, J. & Brus, J. Polymorphic forms of valinomycin investigated by NMR crystallography. *Int. J. Mol. Sci.* **21**, 4907 (2020).
- Du, Y., Frank, D., Chen, Z. X., Struppe, J. & Su, Y. C. Ultrafast magic angle spinning NMR characterization of pharmaceutical solid polymorphism: a posaconazole example. *J. Magn. Reson.* **346**, 107352 (2023).
- Khalaji, M., Paluch, P., Potrzebowski, M. J. & Dudek, M. K. Narrowing down the conformational space with solid-state NMR in crystal structure prediction of linezolid cocrystals. *Solid. State Nucl. Mag.* **121**, 101813 (2022).
- Dudek, M. K. et al. Crystal structure determination of an elusive methanol solvate - hydrate of catechin using crystal structure

- prediction and NMR crystallography. *Crystengcomm* **22**, 4969–4981 (2020).
33. Brus, J. et al. Efficient strategy for determining the atomic-resolution structure of micro- and nanocrystalline solids within polymeric microbeads: domain-edited NMR crystallography. *Macromolecules* **51**, 5364–5374 (2018).
34. Leclaire, J. et al. Structure elucidation of a complex CO<sub>2</sub>-based organic framework material by NMR crystallography. *Chem. Sci.* **7**, 4379–4390 (2016).
35. Holmes, J. B. et al. Imaging active site chemistry and protonation states: NMR crystallography of the tryptophan synthase alpha-aminoacrylate intermediate. *Proc. Natl. Acad. Sci. USA* **119**, e2109235119 (2022).
36. Kumar, A. et al. The atomic-level structure of cementitious calcium silicate hydrate. *J. Phys. Chem. C* **121**, 17188–17196 (2017).
37. Morales-Melgares, A. et al. Atomic-level structure of zinc-modified cementitious calcium silicate hydrate. *J. Am. Chem. Soc.* **144**, 22915–22924 (2022).
38. Kunhi Mohamed, A. et al. The atomic-level structure of cementitious calcium aluminate silicate hydrate. *J. Am. Chem. Soc.* **142**, 11060–11071 (2020).
39. Bamine, T. et al. Understanding local defects in li-ion battery electrodes through combined DFT/NMR studies: application to LiVPO<sub>4</sub>F. *J. Phys. Chem. C* **121**, 3219–3227 (2017).
40. Harper, A. F., Emge, S. P., Magusin, P. C. M. M., Grey, C. P. & Morris, A. J. Modelling amorphous materials via a joint solid-state NMR and X-ray absorption spectroscopy and DFT approach: application to alumina. *Chem. Sci.* **14**, 1155–1167 (2023).
41. Hope, M. A. et al. Nanoscale phase segregation in supramolecular  $\pi$ -templating for hybrid perovskite photovoltaics from NMR crystallography. *J. Am. Chem. Soc.* **143**, 1529–1538 (2021).
42. Cordova, M. et al. Structure determination of an amorphous drug through large-scale NMR predictions. *Nat. Commun.* **12**, 2964 (2021).
43. Nilsson Lill, S. O. et al. Elucidating an amorphous form stabilization mechanism for tenapanor hydrochloride: crystal structure analysis using X-ray diffraction, NMR crystallography, and molecular modeling. *Mol. Pharm.* **15**, 1476–1487 (2018).
44. Kawabata, Y., Wada, K., Nakatani, M., Yamada, S. & Onoue, S. Formulation design for poorly water-soluble drugs based on biopharmaceutics classification system: basic approaches and practical applications. *Int. J. Pharm.* **420**, 1–10 (2011).
45. Babu, N. J. & Nangia, A. Solubility advantage of amorphous drugs and pharmaceutical cocrystals. *Cryst. Growth Des.* **11**, 2662–2679 (2011).
46. Laitinen, R., Löbmann, K., Strachan, C. J., Grohgan, H. & Rades, T. Emerging trends in the stabilization of amorphous drugs. *Int. J. Pharm.* **453**, 65–79 (2013).
47. Yu, L. Amorphous pharmaceutical solids: preparation, characterization and stabilization. *Adv. Drug Deliv. Rev.* **48**, 27–42 (2001).
48. Rossini, A. J. et al. Dynamic nuclear polarization enhanced NMR spectroscopy for pharmaceutical formulations. *J. Am. Chem. Soc.* **136**, 2324–2334 (2014).
49. Rossini, A. J. et al. Dynamic nuclear polarization NMR spectroscopy of microcrystalline solids. *J. Am. Chem. Soc.* **134**, 16899–16908 (2012).
50. Ni, Q. Z. et al. In situ characterization of pharmaceutical formulations by dynamic nuclear polarization enhanced MAS NMR. *J. Phys. Chem. B* **121**, 8132–8141 (2017).
51. Kerber, R. N. et al. Nature and structure of aluminum surface sites grafted on silica from a combination of high-field aluminum-27 solid-state NMR spectroscopy and first-principles calculations. *J. Am. Chem. Soc.* **134**, 6767–6775 (2012).
52. Valla, M. et al. Atomic description of the interface between silica and alumina in aluminosilicates through dynamic nuclear polarization surface-enhanced NMR spectroscopy and first-principles calculations. *J. Am. Chem. Soc.* **137**, 10710–10719 (2015).
53. Lai, J. et al. X-ray and NMR crystallography in an enzyme active site: the indoline quinonoid intermediate in tryptophan synthase. *J. Am. Chem. Soc.* **133**, 4–7 (2011).
54. Klein, A. et al. Atomic-resolution chemical characterization of (2x)72-kDa tryptophan synthase via four- and five-dimensional <sup>1</sup>H-detected solid-state NMR. *Proc. Natl. Acad. Sci.* **119**, e2114690119 (2022).
55. Hartman, J. D., Kudla, R. A., Day, G. M., Mueller, L. J. & Beran, G. J. O. Benchmark fragment-based H-1, C-13, N-15 and O-17 chemical shift predictions in molecular crystals. *Phys. Chem. Chem. Phys.* **18**, 21686–21709 (2016).
56. Hartman, J. D., Monaco, S., Schatschneider, B. & Beran, G. J. O. Fragment-based C-13 nuclear magnetic resonance chemical shift predictions in molecular crystals: an alternative to planewave methods. *J. Chem. Phys.* **143**, 102809 (2015).
57. Joset, K. V. J. & Raghavachari, K. Fragment-based approach for the evaluation of NMR chemical shifts for large biomolecules incorporating the effects of the solvent environment. *J. Chem. Theory Comput.* **13**, 1147–1158 (2017).
58. Gascón, J. A., Sproviero, E. M. & Batista, V. S. QM/MM study of the NMR spectroscopy of the retinyl chromophore in visual rhodopsin. *J. Chem. Theory Comput.* **1**, 674–685 (2005).
59. Jin, X. S., Zhu, T., Zhang, J. Z. H. & He, X. Automated fragmentation QM/MM calculation of NMR chemical shifts for protein-ligand complexes. *Front. Chem.* **6**, 150 (2018).
60. Uluca, B. et al. DNP-enhanced MAS NMR: a tool to snapshot conformational ensembles of alpha-synuclein in different states. *Biophys. J.* **114**, 1614–1623 (2018).
61. Heise, H., Luca, S., de Groot, B. L., Grubmuller, H. & Baldus, M. Probing conformational disorder in neurotensin by two-dimensional solid-state NMR and comparison to molecular dynamics simulations. *Biophys. J.* **89**, 2113–2120 (2005).
62. Siemer, A. B. Advances in studying protein disorder with solid-state NMR. *Solid State Nucl. Magn. Reson* **106**, 101643 (2020).
63. Neal, S., Nip, A. M., Zhang, H. & Wishart, D. S. Rapid and accurate calculation of protein <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N chemical shifts. *J. Biomol. NMR* **26**, 215–240 (2003).
64. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR* **50**, 43–57 (2011).
65. Shen, Y. & Bax, A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR* **48**, 13–22 (2010).
66. Li, J., Bennett, K. C., Liu, Y., Martin, M. V. & Head-Gordon, T. Accurate prediction of chemical shifts for aqueous protein structure on “Real World” data. *Chem. Sci.* **11**, 3180–3191 (2020).
67. Han, Y. et al. Machine learning accelerates quantum mechanics predictions of molecular crystals. *Phys. Rep.* **934**, 1–71 (2021).
68. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
69. Meiler, J. PROSHIFT: protein chemical shift prediction using artificial neural networks. *J. Biomol. NMR* **26**, 25–37 (2003).
70. Gerrard, W. et al. IMPRESSION - prediction of NMR parameters for 3-dimensional chemical structures using machine learning with near quantum chemical accuracy. *Chem. Sci.* **11**, 508–515 (2020).
71. Gerrard, W., Yiu, C. & Butts, C. P. Prediction of N-15 chemical shifts by machine learning. *Magn. Reson. Chem.* **60**, 1087–1092 (2022).
72. Guan, Y., Shree Sowndarya, S. V., Gallegos, L. C., St. John, P. C. & Paton, R. S. Real-time prediction of <sup>1</sup>H and <sup>13</sup>C chemical shifts with

- DFT accuracy using a 3D graph neural network. *Chem. Sci.* **12**, 12012–12026 (2021).
73. Gao, P., Zhang, J., Peng, Q., Zhang, J. & Glezakou, V.-A. General protocol for the accurate prediction of molecular  $^{13}\text{C}/^1\text{H}$  NMR chemical shifts via machine learning augmented DFT. *J. Chem. Inf. Model.* **60**, 3746–3754 (2020).
74. Liu, S. et al. Multiresolution 3D-denseNet for chemical shift prediction in NMR crystallography. *J. Phys. Chem. Lett.* **10**, 4558–4565 (2019).
75. Yang, Z., Chakraborty, M. & White, A. D. Predicting chemical shifts with graph neural networks. *Chem. Sci.* **12**, 10802–10809 (2021).
76. Han, H. & Choi, S. Transfer learning from simulation to experimental data: NMR chemical shift predictions. *J. Phys. Chem. Lett.* **12**, 3662–3668 (2021).
77. Gupta, A., Chakraborty, S. & Ramakrishnan, R. Revving up  $^{13}\text{C}$  NMR shielding predictions across chemical space: benchmarks for atoms-in-molecules kernel machine learning with new data for 134 kilo molecules. *Mach. Learn. Sci. Technol.* **2**, 035010 (2021).
78. Gaumard, R. et al. Regression machine learning models used to predict DFT-computed NMR parameters of zeolites. *Computation* **10**, 74 (2022).
79. Paruzzo, F. M. et al. Chemical shifts in molecular solids by machine learning. *Nat. Commun.* **9**, 4501 (2018).
80. Cordova, M. et al. A machine learning model of chemical shifts for chemically and structurally diverse molecular solids. *J. Phys. Chem. C. Nanomater Interfaces* **126**, 16710–16720 (2022).
81. Kettle, J. G. et al. Discovery of AZD4625, a covalent allosteric inhibitor of the MutantGTPase KRAS(G12C). *J. Med Chem.* **65**, 6940–6952 (2022).
82. Chakraborty, A. et al. AZD4625 is a potent and selective inhibitor of KRASG12C. *Mol. Cancer Ther.* **21**, 1535–1546 (2022).
83. Kragelj, J., Ozenne, V., Blackledge, M. & Jensen, M. R. Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *ChemPhysChem* **14**, 3034–3045 (2013).
84. Nodet, G. et al. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *J. Am. Chem. Soc.* **131**, 17908–17918 (2009).
85. Choy, W.-Y. & Forman-Kay, J. D. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* **308**, 1011–1032 (2001).
86. Filik, J. et al. Processing two-dimensional X-ray diffraction and small-angle scattering data in DAWN 2. *J. Appl. Crystallogr.* **50**, 959–966 (2017).
87. Soper, A. K. & Barney, E. R. Extracting the pair distribution function from white-beam X-ray total scattering data. *J. Appl. Crystallogr.* **44**, 714–726 (2011).
88. Rossini, A. J. et al. Dynamic nuclear polarization surface enhanced NMR spectroscopy. *Acc. Chem. Res.* **46**, 1942–1951 (2013).
89. Lesage, A. et al. Surface enhanced NMR spectroscopy by dynamic nuclear polarization. *J. Am. Chem. Soc.* **132**, 15459–15461 (2010).
90. Sauvée, C. et al. Highly efficient, water-soluble polarizing agents for dynamic nuclear polarization at high frequency. *Angew. Chem. Int. Ed.* **52**, 10858–10861 (2013).
91. Wu, X. L. & Zilm, K. W. Complete spectral editing in CPMAS NMR. *J. Magn. Reson. Ser. A* **102**, 205–213 (1993).
92. Lu, C. et al. OPLS4: improving force field accuracy on challenging regimes of chemical space. *J. Chem. Theory Comput.* **17**, 4291–4300 (2021).
93. (Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021).
94. Bowers, K. J. et al. In: Proceedings of the ACM/IEEE Conference on Supercomputing (SC06).
95. (BIOVIA, Dassault Systèmes, San Diego, 2020).
96. Lin, J. H. Divergence measures based on the Shannon entropy. *IEEE Trans. Inf. Theory* **37**, 145–151 (1991).
97. Hourahine, B. et al. DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *J. Chem. Phys.* **152**, 124101 (2020). **157** (2022).
98. Aradi, B., Hourahine, B. & Frauenheim, T. DFTB+, a sparse matrix-based implementation of the DFTB method. *J. Phys. Chem. A* **111**, 5678–5684 (2007).
99. 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).
100. Rezac, J. Empirical self-consistent correction for the description of hydrogen bonds in DFTB3. *J. Chem. Theory Comput.* **13**, 4804–4817 (2017).
101. Yang, Y., Yu, H. B., York, D., Cui, Q. & Elstner, M. Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction. *J. Phys. Chem. A* **111**, 10861–10873 (2007).
102. Gaus, M., Goez, A. & Elstner, M. Parametrization and benchmark of DFTB3 for organic molecules. *J. Chem. Theory Comput.* **9**, 338–354 (2013).
103. 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).

## Acknowledgements

This work was supported by AstraZeneca, Swiss National Science Foundation Grant No. 200020\_212046, and by the NCCR MARVEL.

## Author contributions

P.M. performed the solid-state NMR experiments. S.N.L. performed the MD simulations. M.C. computed chemical shifts on MD structures, performed the structural analysis, visualization, and scoring, and computed intermolecular complex formation energies. A.C., M.K., J.McC., A.S.A., A.C.P., and S.T.N. prepared and chemically characterized the samples in solid and solution forms. M.C., P.M., S.N.L., S.T.N., J.McC., S.S., and L.E. analyzed the results. S.S. and L.E. conceived and supervised the research. M.C., P.M., 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/s41467-023-40853-2>.

**Correspondence** and requests for materials should be addressed to Staffan Schantz or Lyndon Emsley.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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) 2023