

Supporting information for

# Combining double-hybrid functionals with rSCAN yields solid-state $^{13}\text{C}$ chemical shifts with sub-ppm accuracy

Harry Brough,<sup>†</sup> Chris W. Cook,<sup>†</sup> John M. Griffin,<sup>†</sup> and Michael J. G. Peach<sup>†</sup>

<sup>†</sup> Department of Chemistry, Lancaster University, Lancaster, LA1 4YB, United Kingdom

## Contents

<b>1</b>	<b>Sets of Crystals</b>	<b>2</b>
<b>2</b>	<b>Example of Applying a Correction</b>	<b>4</b>
<b>3</b>	<b>Comparison of Computational Cost</b>	<b>5</b>
<b>4</b>	<b>Effect of Geometry Optimisation and GIPAW Method on Chemical Shifts</b>	<b>6</b>
<b>5</b>	<b>Sensitivity of Corrections to Basis Set</b>	<b>7</b>
<b>6</b>	<b>Dispersion and the SCAN Functional</b>	<b>9</b>
<b>7</b>	<b>Implicit Solvation Models</b>	<b>11</b>
<b>8</b>	<b>Cluster Corrections</b>	<b>13</b>
<b>9</b>	<b>Outliers</b>	<b>14</b>
<b>10</b>	<b>Experimental Spectra</b>	<b>15</b>
	<b>References</b>	<b>23</b>

## 1 Sets of Crystals

The organic crystals used in our test set and corresponding Cambridge Structural Database<sup>1</sup> (CSD) reference codes and diffraction methods are shown in Table S1. Similarly, Table S2 shows this information for the previously reported test set<sup>2</sup> on which we validated the performance of corrections with mPW2PLYP-CPCM.

**Table S1** Molecular crystals, CSD reference codes and diffraction methods for our <sup>13</sup>C test set.

Number	Molecular Crystal	Reference	Diffraction
1	L-Alanine	LALNIN12	Neutron
2	L-Ascorbic Acid	LASCAC10	Neutron
3	Aspirin	ACSALA	X-Ray
4	L-Cysteine	LCYSTN21	X-Ray
5	Cytosine	CYTSIN	X-Ray
6	$\alpha$ -D-Glucose	GLUCSA11	Neutron
7	Paracetamol	HXACAN26	X-Ray
8	2-Pyridone	PYRIDO11	Neutron
9	L-Serine	LSERIN01	X-Ray
10	L-Threonine	LTHREO01	Neutron
11	Thymine	THYMIN01	X-Ray
12	L-Tyrosine	LTYROS11	Neutron
13	Uracil	URACIL	X-Ray

**Table S2** Molecular crystals, CSD reference codes and diffraction methods for the  $^{13}\text{C}$  test set on which mPW2PLYP–CPCM corrections were validated.<sup>2</sup>

<b>Molecular Crystal</b>	<b>Reference</b>	<b>Diffraction</b>
Adenosine	ADENOS12	X-Ray
L-Asparagine monohydrate	ASPARM03	Neutron
$\beta$ -D-Fructopyranose	FRUCTO02	Neutron
L-Glutamine	GLUTAM01	Neutron
Glycine	GLYCIN29	X-Ray
Paracetamol	HXACAN26	X-Ray
L-Alanine	LALNIN12	Neutron
L-Cysteine	LCYTSN21	X-Ray
L-Serine	LSERIN01	X-Ray
L-Serine monohydrate	LSERMH10	Neutron
L-Threonine	LTHREO01	Neutron
L-Tyrosine	LYTROS11	Neutron
Methyl- $\beta$ -D-galactopyranoside	MBDGAL02	Neutron
Methyl- $\alpha$ -D-manopyranoside	MEMANP11	Neutron
Methyl- $\alpha$ -D-galactopyranoside monohydrate	MGALPY01	Neutron
Methyl- $\alpha$ -D-glucopyranoside	MGLUCP11	Neutron
Pentaerythrytol	PERYTO10	Neutron
$\alpha$ -L-Rhamnose monohydrate	RHAMAH12	Neutron
Sucrose	SUCROS04	Neutron
4-Aminobenzenesulfonamide	SULAMD06	X-Ray
Triphenylene	TRIPHE11	Neutron

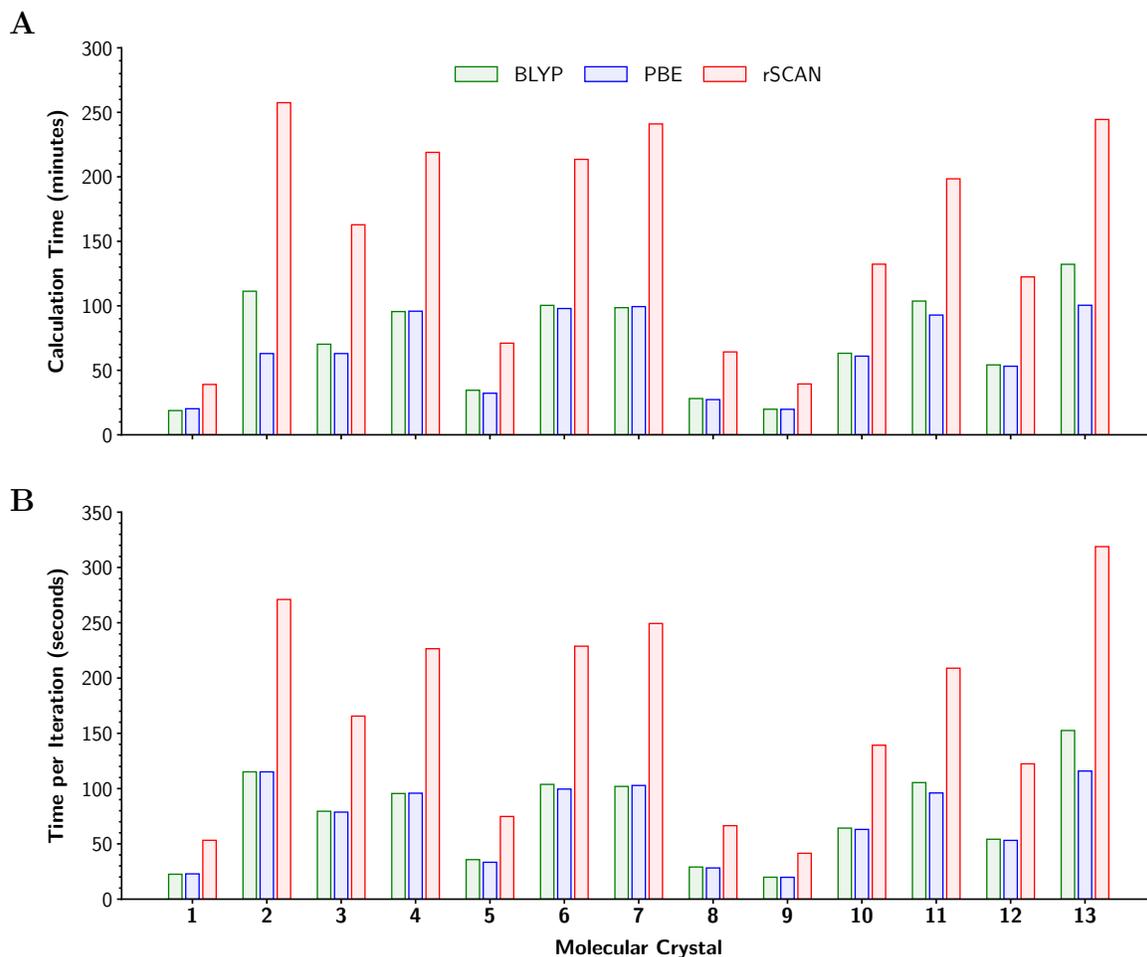
## 2 Example of Applying a Correction

This section describes the step by step process to apply a monomer correction to improve the calculated chemical shifts of an organic crystal. In this example, a mPW2LPYP–CPCM correction is applied to PBE GIPAW shieldings.

- Obtain a crystal structure, perhaps from the Crystal Structure Database.<sup>1</sup>
- Fully optimise the crystal structure in a plane wave DFT code such as CASTEP, using an exchange–correlation functional of your choice — we recommend rSCAN.
- On the optimised crystal structure, run a GIPAW calculation to calculate shieldings with PBE.
- Use molecular visualisation software such as GaussView to extract a single molecular structure from the optimised geometry.
- On this isolated molecule, run shielding calculations in a gauge-including atomic orbital basis with PBE and mPW2PLYP. Perform both calculations in a CPCM with a high dielectric constant ( $\epsilon > 20$ ).
- Calculate the corrected set of shieldings by  $\sigma = \sigma_{\text{crystal}}^{\text{PBE}} - \sigma_{\text{molecule}}^{\text{PBE-CPCM}} + \sigma_{\text{molecule}}^{\text{mPW2PLYP-CPCM}}$ .
- Partially or fully assign your experimental solid-state NMR spectrum by matching the largest chemical shifts with the smallest shieldings.
- Run a linear regression to determine the gradient,  $m$ , which should be approximately  $-1$ , and the intercept  $c$ , which is the shielding of the chemical shift reference.
- Convert the corrected shieldings to calculated chemical shifts by  $\delta_{\text{calc}} = m\sigma + c$ .
- Compare these calculated shifts to experimental chemical shifts.
- Optionally repeat these steps for another potential structure (ie. a polymorph), and compare the mean error in the calculated shifts to determine which is the better structure.

### 3 Comparison of Computational Cost

Figure S1 shows that periodic optimisations with BLYP and PBE generally took similar amounts of computational time, whereas rSCAN calculations took on average  $2.5\times$  longer than PBE. There was no significant difference between the relative overall calculation times and relative times to calculate the energy at each geometry iteration, so rSCAN is slower to evaluate the energy but does not require more geometry iterations. Shielding calculations with rSCAN also took approximately  $2.5\times$  longer than those with PBE.



**Figure S1** Total calculation time (**A**) and calculation time per geometry iteration (**B**) of CASTEP geometry optimisations of molecular crystals with PBE, BLYP and rSCAN. Calculations run on 120 parallel processes using Intel(R) Xeon(R) Gold 6248 @ 2.50GHz processors with 4 GB memory per core.

## 4 Effect of Geometry Optimisation and GIPAW Method on Chemical Shifts

**Table S3** Comparison of RMSE, MAE, IQR, and MAX values for different underlying geometries and subsequent GIPAW calculations. In geometries labelled “H Only” only hydrogen atom positions were relaxed. All values in ppm.

<b>Geometry</b>	<b>GIPAW</b>	<b>RMSE</b>	<b>MAE</b>	<b>IQR</b>	<b>MAX</b>
Unoptimised	PBE	4.4	3.8	7.1	12.5
	rSCAN	4.2	3.5	6.2	12.8
PBE H Only	PBE	2.5	2.1	3.6	7.3
	rSCAN	2.6	2.1	3.8	7.9
BLYP H Only	PBE	2.5	2.1	3.6	7.2
	rSCAN	2.6	2.1	3.7	7.8
rSCAN H Only	PBE	2.5	2.1	3.5	7.3
	rSCAN	2.6	2.1	3.6	7.9
PBE All Atom	PBE	2.1	1.7	2.4	5.5
	rSCAN	2.4	2.0	3.7	6.1
BLYP All Atom	PBE	2.3	1.8	2.6	5.6
	rSCAN	2.7	2.3	4.2	6.3
rSCAN All Atom	PBE	2.1	1.6	2.6	4.7
	rSCAN	2.3	1.9	3.9	5.8

## 5 Sensitivity of Corrections to Basis Set

To confirm our set of molecules was insensitive to the choice of basis for the isolated molecular calculations, a series of corrections were calculated to the PBE shieldings on the rSCAN geometries with PBE0, RI-MP2 and r<sup>2</sup>SCAN-D, with error statistics shown in Tables S4, S5 and S6 respectively.

The basis sets tested included the widely used correlation-consistent  $N$ -tuple-zeta sets by Dunning, cc-pVNZ,<sup>3</sup> as well as core-valence weighted basis sets in the same style, cc-pwCVNZ,<sup>4</sup> and sets with additional diffuse functions, aug-cc-pVNZ.<sup>5</sup> In addition, a basis set with both diffuse and core-valence weighted functions, aug-cc-pwCVDZ was tested, along with the polarisation-consistent basis sets designed for nuclear shielding calculation by Jensen, pcS-1 and pcS-2.<sup>6</sup>

**Table S4** Basis set dependence of error statistics (ppm) for PBE0 correction to PBE GIPAW on structures optimised with rSCAN.

Basis Set	RMSE	MAE	MAX	IQR
cc-pVDZ	1.6	1.3	4.3	2.1
cc-pVTZ	1.5	1.2	4.3	2.0
cc-pVQZ	1.5	1.2	4.3	2.0
cc-pwCVDZ	1.6	1.3	4.3	2.1
cc-pwCVTZ	1.5	1.2	4.3	2.1
aug-cc-pVDZ	1.6	1.3	4.3	2.1
aug-cc-pVTZ	1.5	1.2	4.4	2.0
aug-cc-pwCVDZ	1.5	1.2	4.3	2.0
pcS-1	1.6	1.3	4.3	2.3
pcS-2	1.5	1.2	4.3	2.1

**Table S5** Basis set dependence of error statistics (ppm) for RI-MP2 correction to PBE GIPAW on structures optimised with rSCAN.

<b>Basis Set</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAX</b>	<b>IQR</b>
cc-pVDZ	2.1	1.8	4.7	2.8
cc-pVTZ	1.8	1.5	3.5	2.6
cc-pVQZ	1.8	1.5	3.7	2.6
cc-pwCVDZ	2.1	1.7	4.7	2.9
cc-pwCVTZ	1.8	1.5	4.0	2.6
aug-cc-pVDZ	1.9	1.7	4.0	3.1
aug-cc-pVTZ	1.8	1.5	3.5	2.4
aug-cc-pwCVDZ	2.0	1.5	6.1	2.3
pcS-1	1.9	1.6	4.4	2.6
pcS-2	1.8	1.5	3.7	2.7

**Table S6** Basis set dependence of error statistics (ppm) for r<sup>2</sup>SCAN correction to PBE GIPAW on structures optimised with rSCAN.

<b>Basis Set</b>	<b>RMSE</b>	<b>MAE</b>	<b>MAX</b>	<b>IQR</b>
cc-pVDZ	2.1	1.7	5.3	3.3
cc-pVTZ	2.0	1.7	5.5	3.3
cc-pVQZ	2.0	1.7	5.4	3.1
cc-pwCVDZ	2.0	1.6	5.3	3.1
cc-pwCVTZ	2.0	1.7	5.4	2.6
aug-cc-pVDZ	2.2	1.7	4.0	3.2
aug-cc-pVTZ	2.1	1.7	5.5	3.3
aug-cc-pwCVDZ	2.1	1.7	5.5	3.2
pcS-1	1.9	1.6	5.4	2.9
pcS-2	2.1	1.7	5.5	3.1

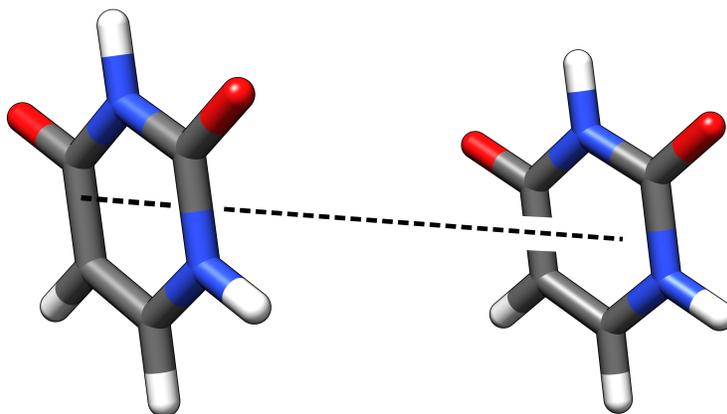
## 6 Dispersion and the SCAN Functional

Because phase stability in ice polymorphs is highly dependent on dispersion, the authors of SCAN suggest its success at predicting volume changes in these systems<sup>7</sup> is due to its “ability to capture the intermediate-range van der Waals interaction” against the conventional wisdom that mGGAs cannot account for this effect.<sup>8</sup>

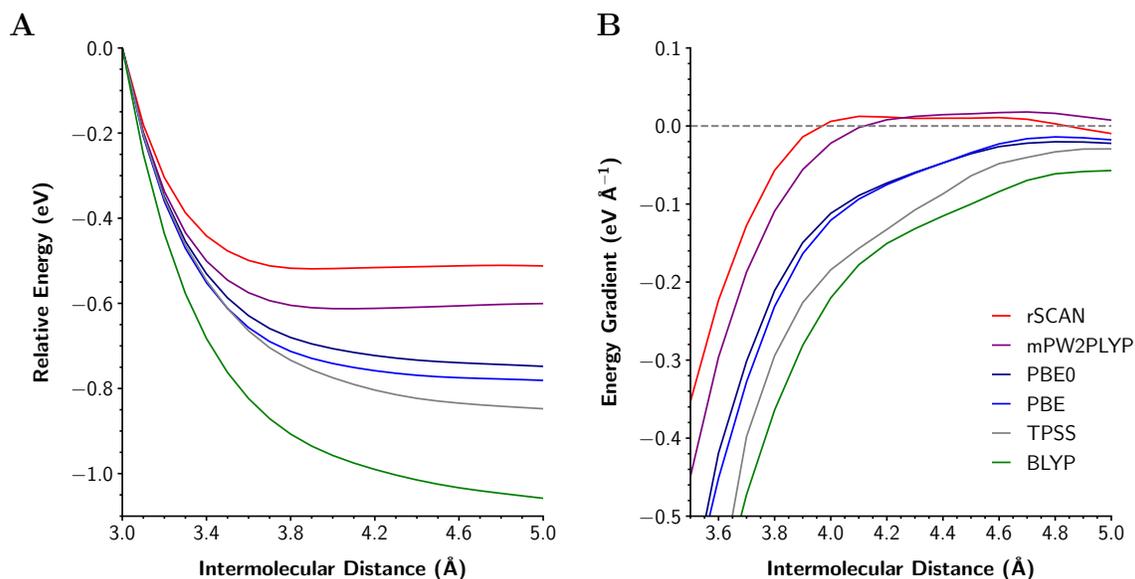
To investigate whether SCAN considers dispersion interactions, we performed energy calculations as a function of distance over the interfacial bond in a uracil dimer, shown in Figure S2.

Energies were calculated in the aug-cc-pVTZ basis with the GGA functionals PBE<sup>9</sup> and BLYP,<sup>10,11</sup> the mGGA functionals TPSS,<sup>12</sup> SCAN,<sup>8</sup> rSCAN<sup>13</sup> and r<sup>2</sup>SCAN,<sup>14</sup> the hybrid functionals PBE0,<sup>15</sup> B3LYP<sup>16</sup> and CAM-B3LYP,<sup>17</sup> and the double-hybrid functionals B2PLYP<sup>18</sup> and mPW2PLYP.<sup>19</sup> Double-hybrids incorporate non-local MP2 correlation energy so are expected to detect weak interactions.<sup>20</sup>

Only the SCAN family and the double-hybrids bound the uracil dimer weakly in the 3–5 Å range, demonstrated by rSCAN and mPW2PLYP in Figure S3. Upon addition of Grimme’s D4 semi-empirical dispersion correction scheme,<sup>21</sup> most methods gave an optimal binding distance of 3.8 Å, as shown in Table S7. The binding distances of the SCAN family of functionals were much closer to this value when uncorrected than the double-hybrids tested, with the regularisation slightly reducing agreement with the D4 optimal distance.



**Figure S2** Structure of uracil dimer and vector over which energies were calculated shown by the dashed black line. Carbon atoms shown in grey, hydrogens in white, oxygens in red and nitrogens in blue.



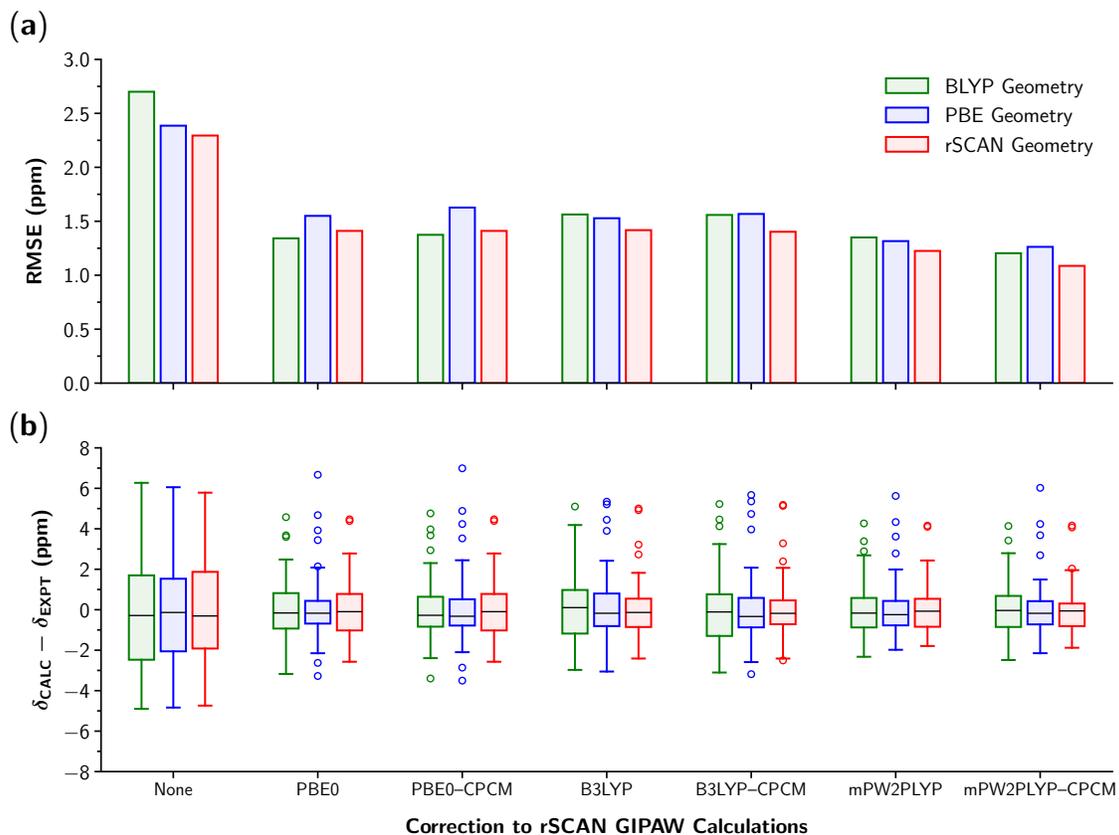
**Figure S3** Energy change from energy at 3 Å (**A**) and energy gradient (**B**) in uracil dimer as a function of distance, with data points every 0.1 Å, for DFAs without dispersion correction. The DFAs that bind the uracil dimer in the 3–5 Å range cross the dashed line at an energy gradient of zero.

**Table S7** Energy minimum distances in the uracil dimer from DFAs with and without D4 correction.

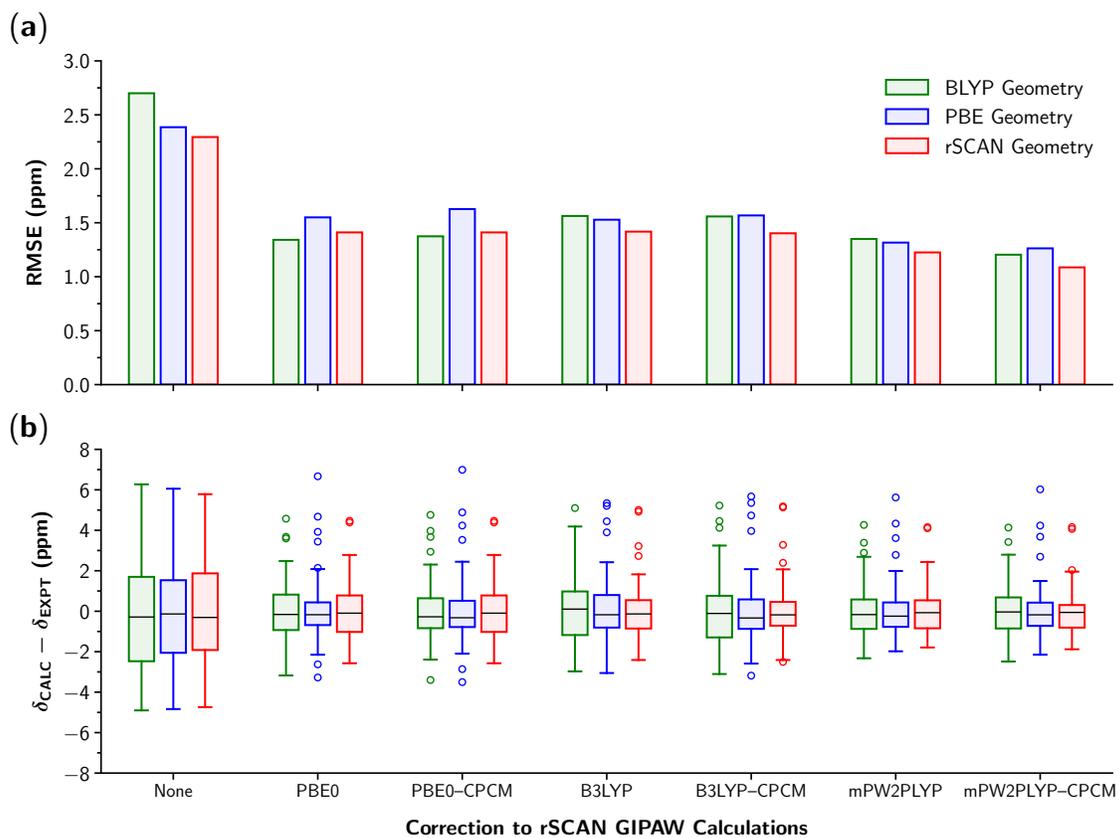
Density-functional Approximation	Binding Distance with No Correction / Å	Binding Distance with D4 Correction / Å
PBE	—	3.9
BLYP	—	3.7
TPSS	—	3.8
PBE0	—	3.8
B3LYP	—	3.9
CAM-B3LYP	—	3.8
B2PLYP	4.5	3.8
mPW2PLYP	4.1	3.8
SCAN	3.9	3.8
rSCAN	4.0	3.8
r <sup>2</sup> SCAN	4.0	3.8

## 7 Implicit Solvation Models

Figures S4 and S5 show the narrowing of the error distribution that arises from applying a CPCM correction to PBE and rSCAN GIPAW respectively, compared to conventional hybrid and double-hybrid vacuum monomer corrections.



**Figure S4** Root-mean-square errors (a) and error distribution (b) of various monomer correction approaches to the PBE periodic shielding calculations, on geometries optimised with BLYP, PBE and rSCAN. Corrections with mGGA shieldings were calculated with  $\tau_{\text{D}}(\mathbf{r})$ .



**Figure S5** Root-mean-square errors (a) and error distribution (b) of various monomer correction approaches to the rSCAN periodic shielding calculations, on geometries optimised with BLYP, PBE and rSCAN. The corrections to rSCAN shown were calculated by subtracting rSCAN  $\tau_{\text{MS}}(\mathbf{r})$  shieldings and adding mGGA  $\tau_{\text{D}}(\mathbf{r})$  shieldings.

## 8 Cluster Corrections

To cover the central molecule, different crystal structures required different numbers of surrounding molecules. The number of individual molecules included in each cluster is shown in Table S8.

**Table S8** Number of molecules used in clusters for organic crystals used in our  $^{13}\text{C}$  test set.

Number	Molecular Crystal	Number of Molecules in Cluster
1	L-Alanine	9
2	L-Ascorbic Acid	10*
3	Aspirin	10
4	L-Cysteine	10
5	Cytosine	13
6	$\alpha$ -D-Glucose	11
7	Paracetamol	11
8	2-Pyridone	15
9	L-Serine	13
10	L-Threonine	13
11	Thymine	13
12	L-Tyrosine	9
13	Uracil	9

\* The L-ascorbic acid calculations were done with a separate cluster for each molecule in the asymmetric unit, each with 10 molecules, to reduce computational cost.

## 9 Outliers

In the box and whiskers plots presented in the article, chemical shifts more than  $1.5\times$  the interquartile range away from the mean are shown by circles. In this section, these outliers are discussed. It is important to note that the presence of more outliers does not indicate an inferior method — in fact, a more reliable method will have a tighter interquartile range, so the definition of an outlier will include more data points. Investigation of outliers can be helpful, however, to determine if certain environments are consistently poorly described.

For the hydrogen-only optimisations in Figure 2, there is one outlier for each set of calculations. This calculated chemical shift, with an absolute deviation around 7 ppm, is the O–C–O carbon in  $\alpha$ -D-glucose. Upon optimisation of heavy atoms, the deviation falls to around 4 ppm, and is then essentially eliminated by monomer corrections, to below 1 ppm in error. The PBE GIPAW calculations on the PBE and BLYP geometries have two outliers each. One of these (for both cases) is the cytosine N–C–N peak. A correction with PBE0 halves these errors to around 4 ppm, while a correction with mPW2PLYP reduces the error to around 2 ppm. The other outlier for the PBE geometry is one of the challenging ascorbic acid environments discussed in the article, and the other outlier for the BLYP geometry is the O–C–O peak in  $\alpha$ -D-glucose.

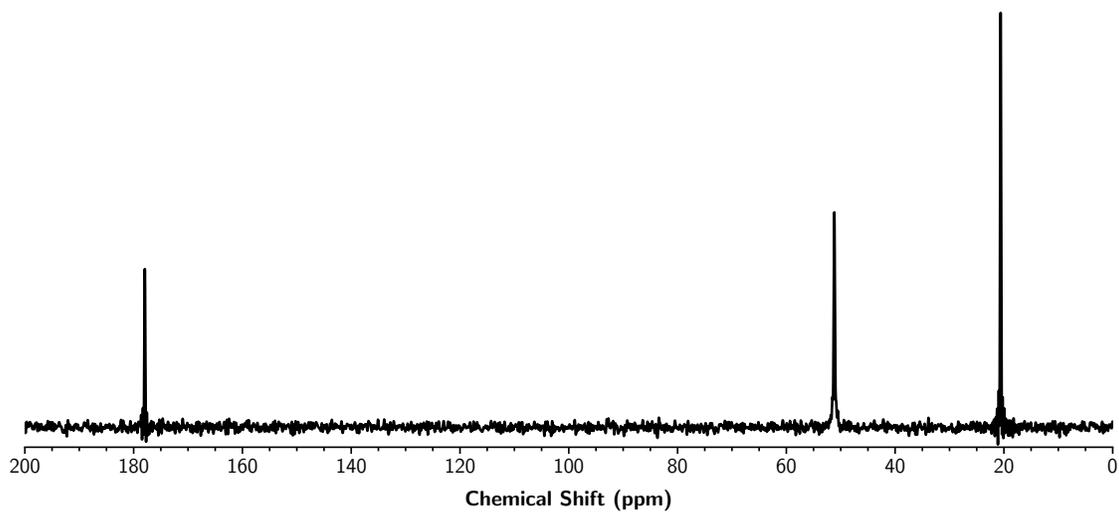
In Figure 8, which shows a comparison of corrected and uncorrected methods to calculate chemical shifts, PBE0 correction to rSCAN GIPAW on the PBE geometries has several outliers. These include both challenging ascorbic acid environments, the aspirin O–C–O peak, the cytosine N–C–N and N–C–O peaks and the N–C–O peaks in 2-pyridone. On the same set of geometries, correction with mPW2PLYP–CPCM rather than PBE0 yields four outliers: the two difficult ascorbic acid peaks and the aspirin O–C–O and (HC)–C–O peaks.

Overall, carbon environments with two adjacent heteroatoms seem particularly sensitive to the treatment of electronic structure and geometry. Therefore, an organic crystal with several of these environments may present a more difficult case for conventional NMR crystallography methods, such as optimisation and GIPAW calculation with PBE, and might benefit substantially from higher quality geometries and monomer corrections.

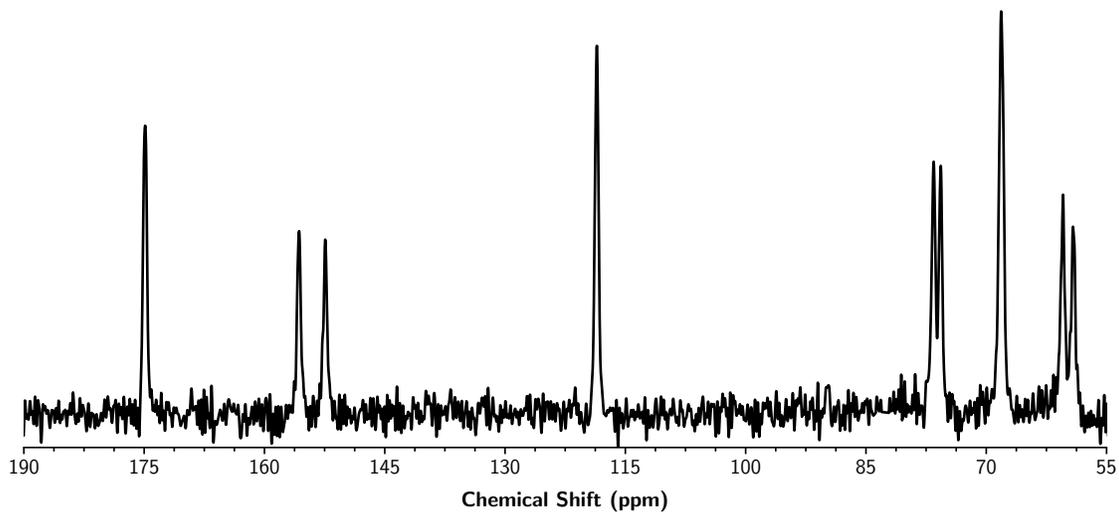
## 10 Experimental Spectra

This section includes the unassigned cross-polarisation magic-angle spinning solid-state  $^{13}\text{C}$  NMR spectra for our set of molecules, in Figures S6–S18. Assignments of the peaks can be found in the accompanying Excel spreadsheet.

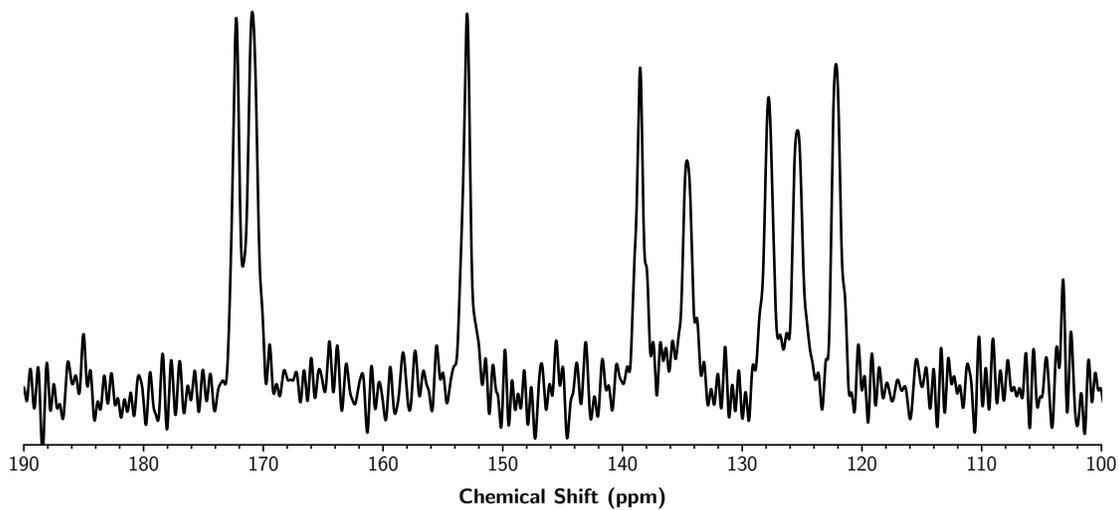
Chemical shifts were referenced to tetramethylsilane with the methyl group of L-alanine at 20.5 ppm used as a secondary reference.



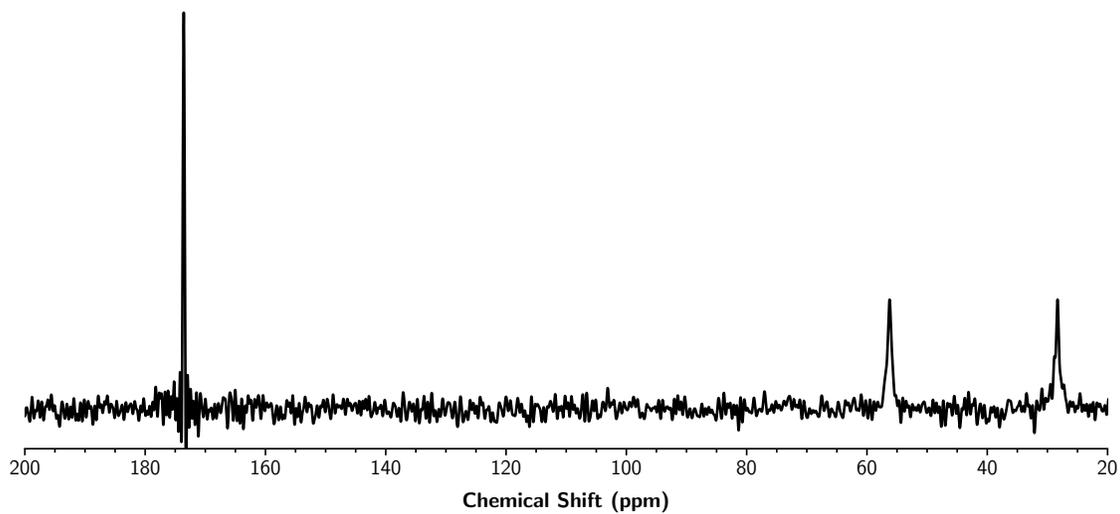
**Figure S6** Solid-state  $^{13}\text{C}$  NMR spectrum of L-alanine.



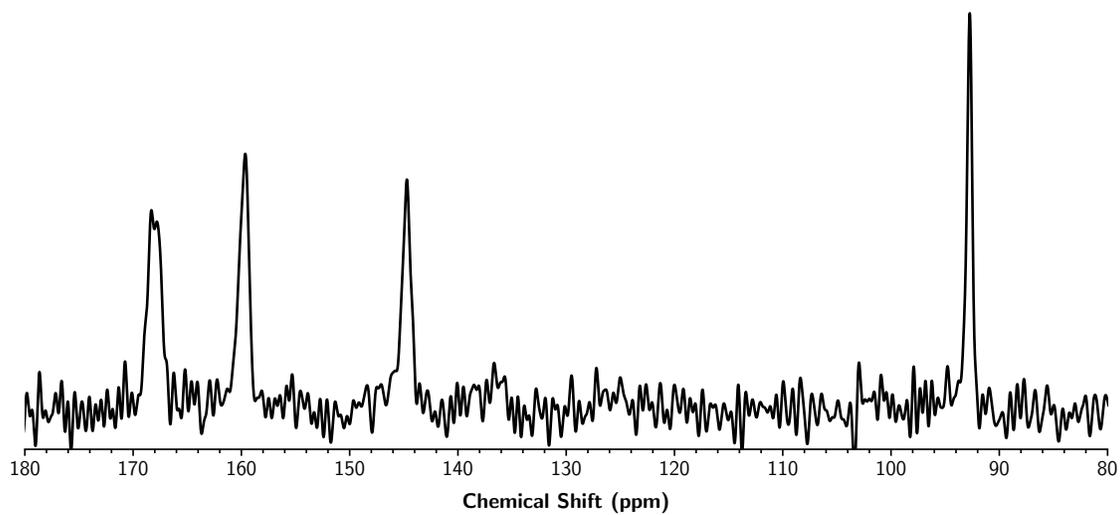
**Figure S7** Solid-state  $^{13}\text{C}$  NMR spectrum of L-ascorbic acid.



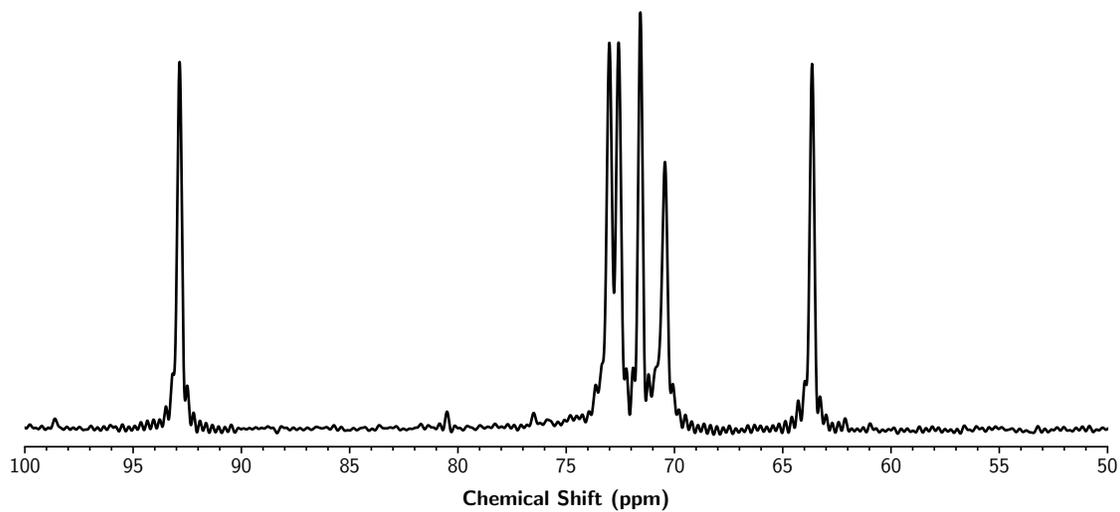
**Figure S8** Solid-state  $^{13}\text{C}$  NMR spectrum of aspirin.



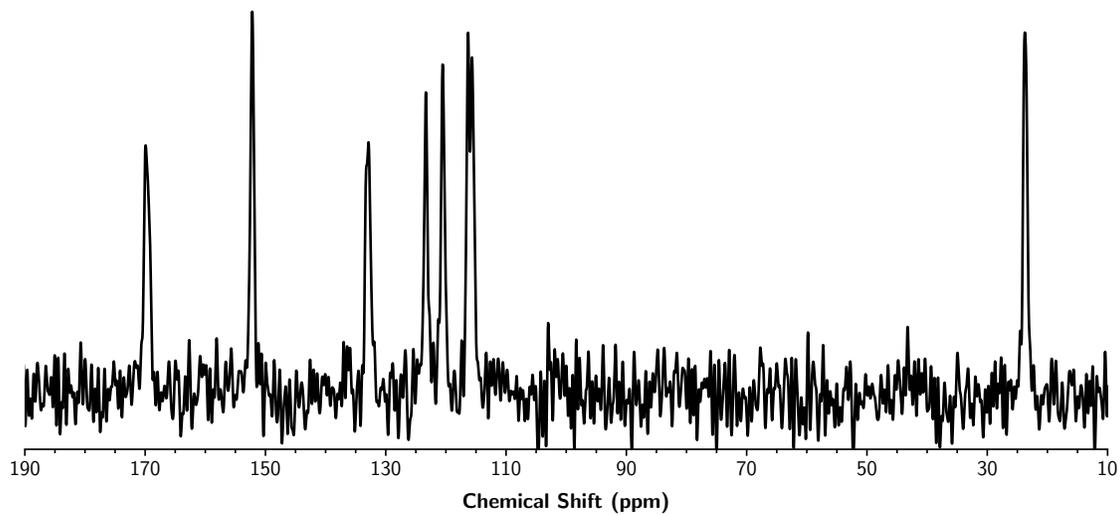
**Figure S9** Solid-state  $^{13}\text{C}$  NMR spectrum of L-cysteine.



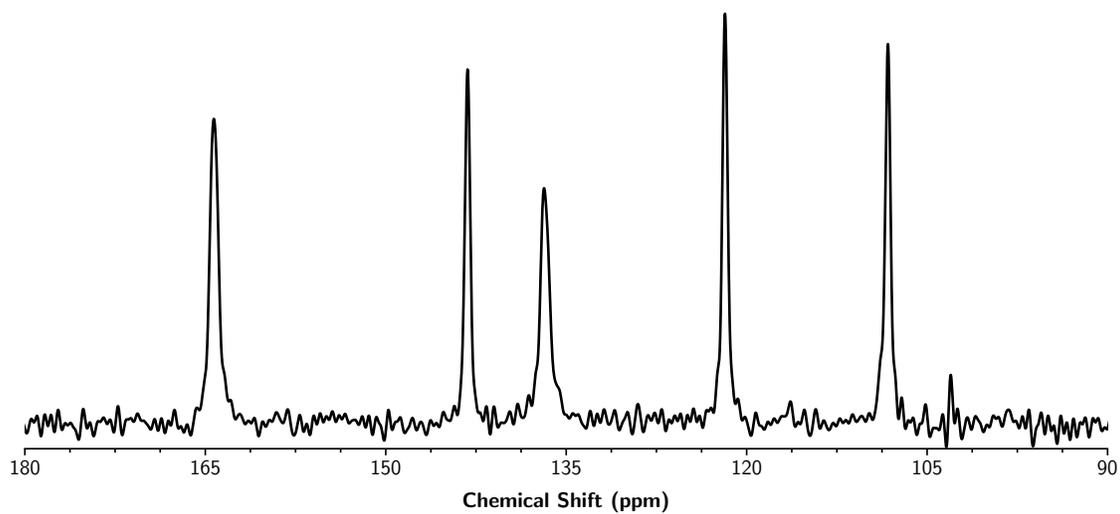
**Figure S10** Solid-state  $^{13}\text{C}$  NMR spectrum of cytosine.



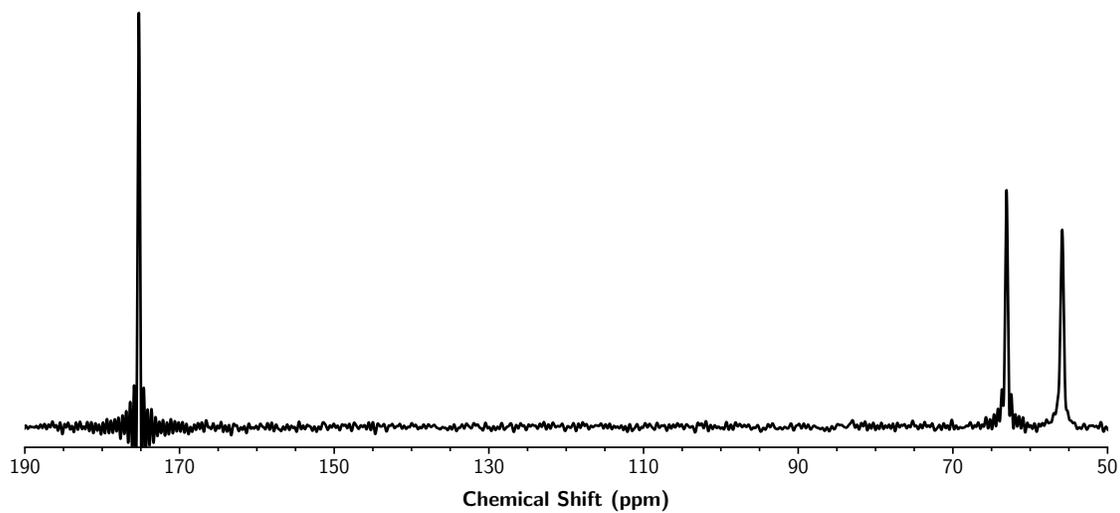
**Figure S11** Solid-state  $^{13}\text{C}$  NMR spectrum of  $\alpha$ -D-glucose.



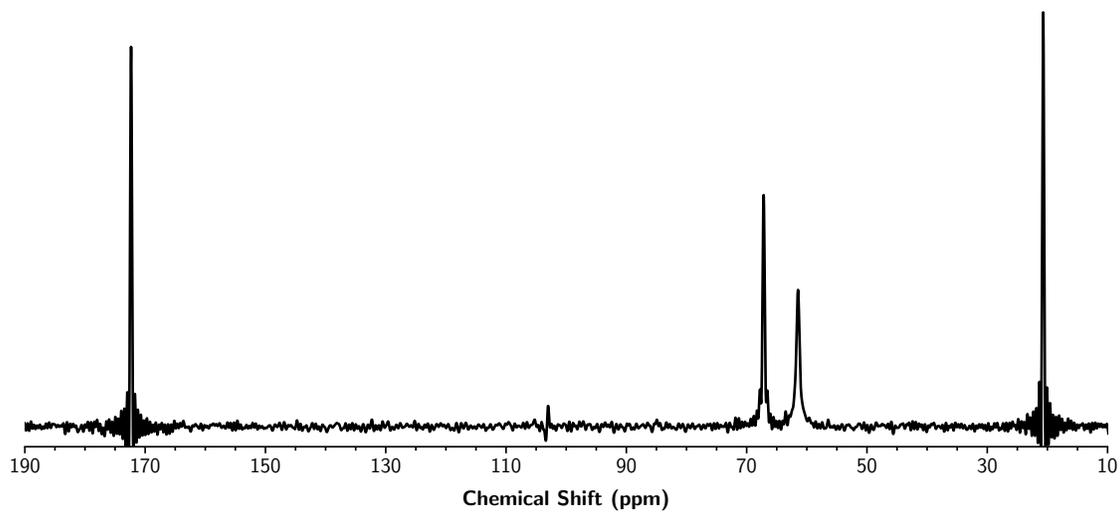
**Figure S12** Solid-state  $^{13}\text{C}$  NMR spectrum of paracetamol.



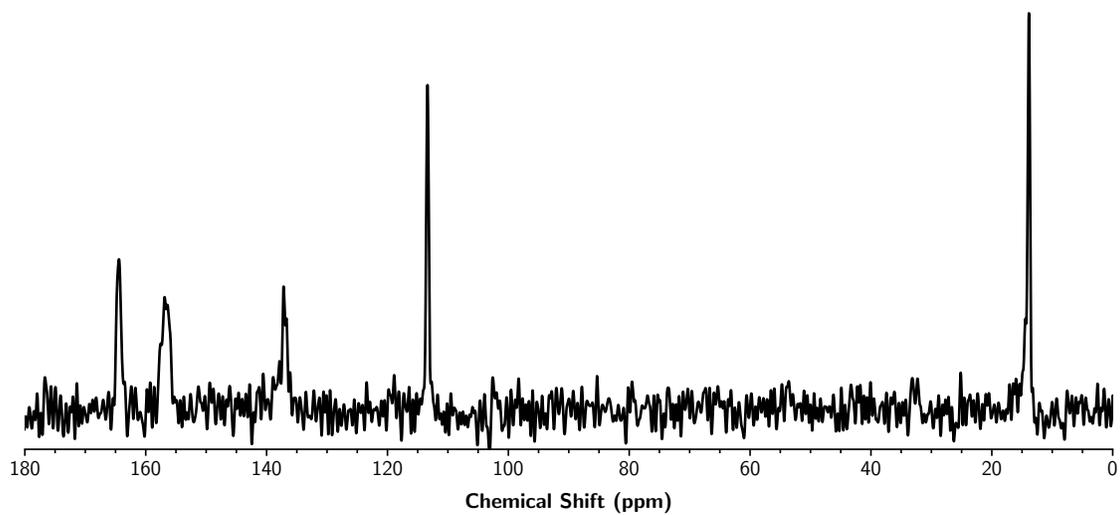
**Figure S13** Solid-state  $^{13}\text{C}$  NMR spectrum of 2-pyridone.



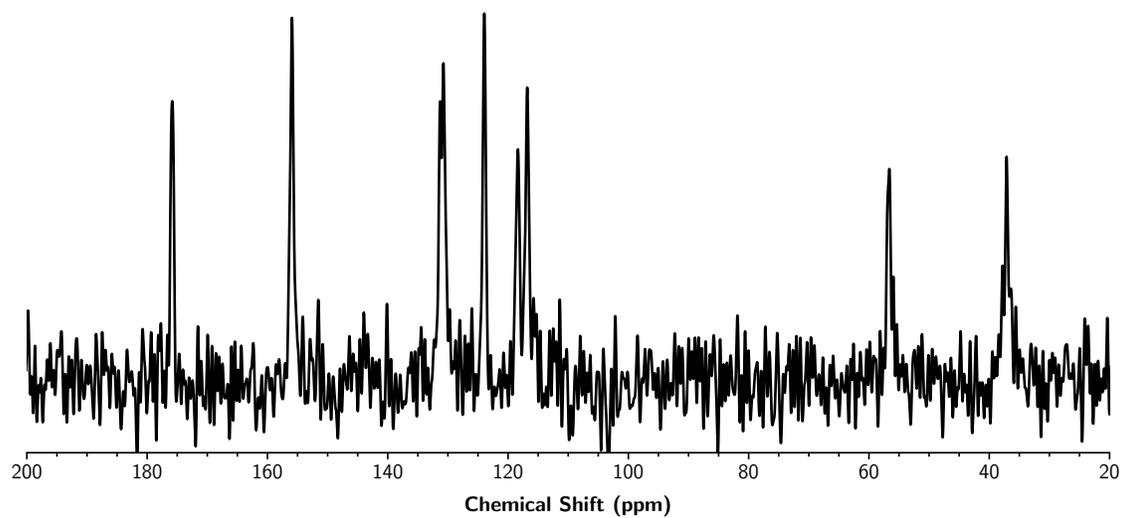
**Figure S14** Solid-state  $^{13}\text{C}$  NMR spectrum of L-serine.



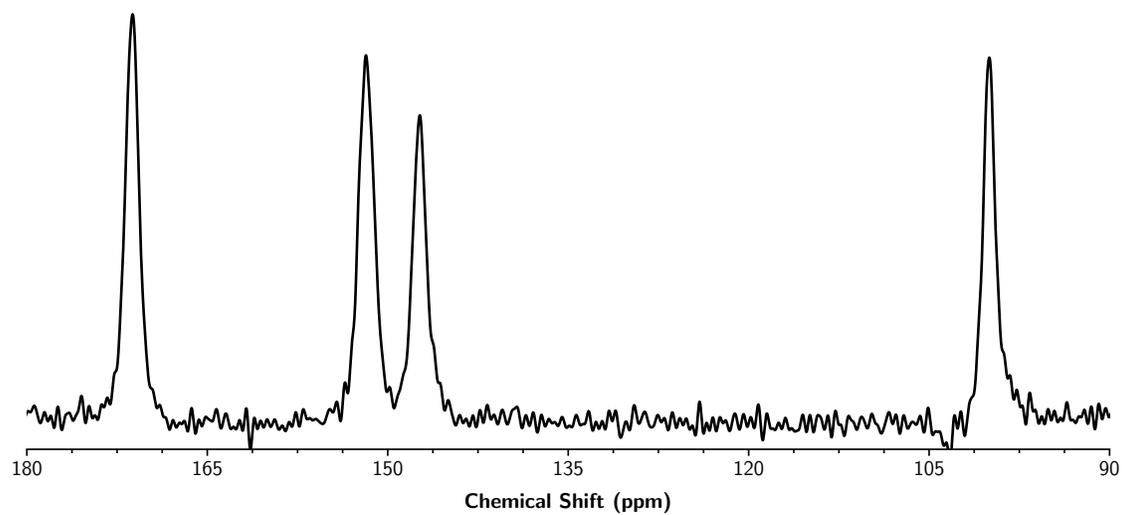
**Figure S15** Solid-state  $^{13}\text{C}$  NMR spectrum of L-threonine.



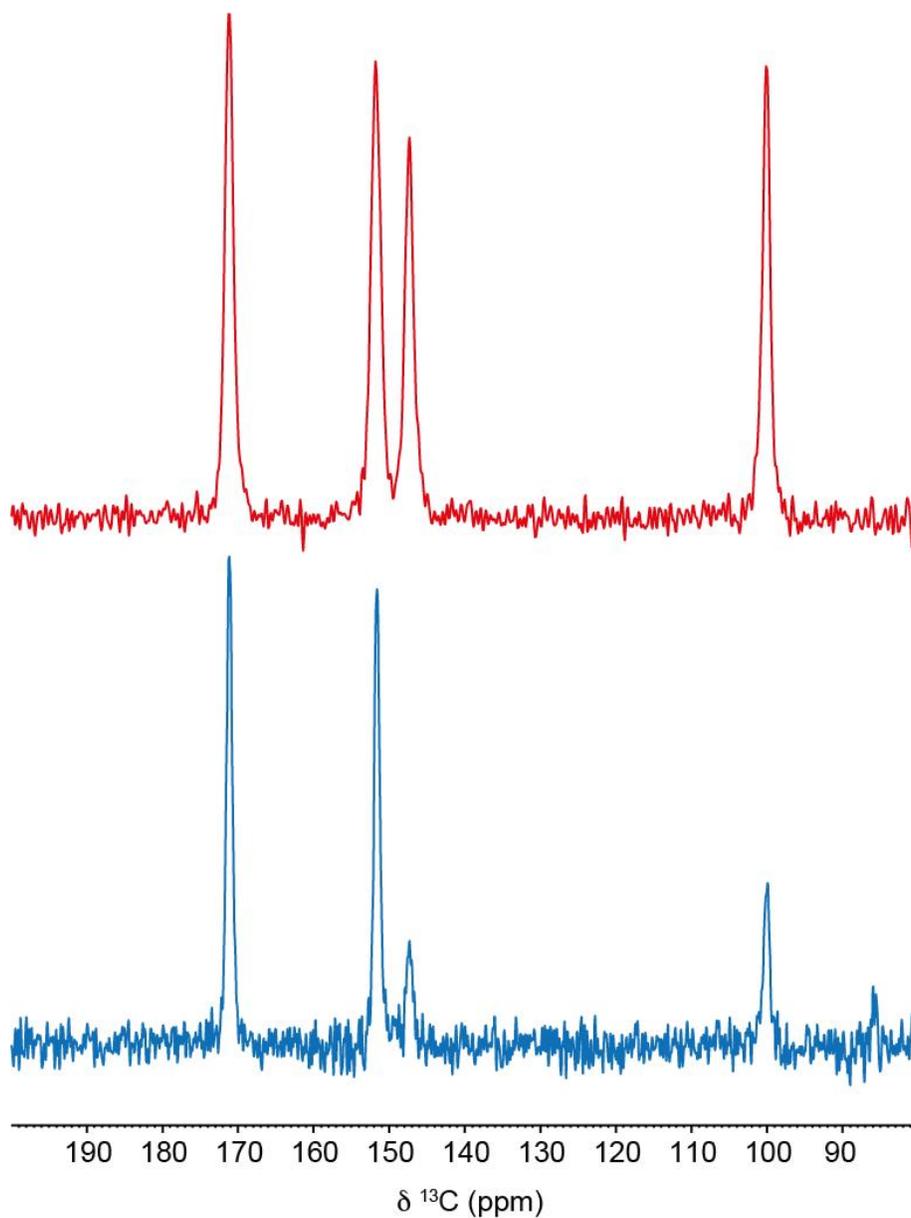
**Figure S16** Solid-state  $^{13}\text{C}$  NMR spectrum of thymine.



**Figure S17** Solid-state  $^{13}\text{C}$  NMR spectrum of L-tyrosine.



**Figure S18** Solid-state  $^{13}\text{C}$  NMR spectrum of uracil.



**Figure S19**  $^{13}\text{C}$  CP MAS NMR spectrum (red) compared with a dipolar dephasing spectrum (blue) of uracil. The dipolar dephasing spectrum was recorded with a dephasing time of 380  $\mu\text{s}$ , during which  $^1\text{H}$  decoupling was not applied. This selectively dephases carbon environments in close proximity to dipolar-coupled  $^1\text{H}$ , i.e., those which are directly bonded. In the dipolar dephasing spectrum, the relative intensities of resonances at 147.2 and 100.0 ppm are significantly reduced, confirming that these resonances correspond to C-H carbons within the uracil structure.

## References

- (1) C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. B*, 2016, **72**, 171–179.
- (2) R. J. Iuliucci, J. D. Hartman and G. J. O. Beran, *J. Phys. Chem. A*, 2023, **127**, 2846–2858.
- (3) J. Dunning, Thom H., *J. Chem. Phys.*, 1989, **90**, 1007–1023.
- (4) K. A. Peterson and J. Dunning, Thom H., *J. Chem. Phys.*, 2002, **117**, 10548–10560.
- (5) R. A. Kendall, J. Dunning, Thom H. and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.
- (6) F. Jensen, *J. Chem. Theory Comput.*, 2008, **4**, 719–727.
- (7) P. M. Piaggi, A. Z. Panagiotopoulos, P. G. Debenedetti and R. Car, *J. Chem. Theory Comput.*, 2021, **17**, 3065–3077.
- (8) J. Sun, A. Ruzsinszky and J. P. Perdew, *Phys. Rev. Lett.*, 2015, **115**, 036402.
- (9) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- (10) A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- (11) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.
- (12) J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.
- (13) A. P. Bartók and J. R. Yates, *J. Chem. Phys.*, 2019, **150**, 161101.
- (14) J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew and J. Sun, *J. Chem. Phys. Lett.*, 2020, **11**, 8208–8215.
- (15) C. Adamo, M. Cossi and V. Barone, *J. Mol. Struct.*, 1999, **493**, 145–157.
- (16) A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.
- (17) T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51–57.
- (18) S. Grimme and F. Neese, *J. Chem. Phys.*, 2007, **127**, 154116.
- (19) T. Schwabe and S. Grimme, *Phys. Chem. Chem. Phys.*, 2006, **8**, 4398–4401.
- (20) M. Bursch, J.-M. Mewes, A. Hansen and S. Grimme, *Angew. Chem. Int. Edit.*, 2022, **61**, e202205735.
- (21) E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2019, **150**, 154122.