Supporting Information for "Reliable fragment-based ^{1}H, ^{13}C, ^{15}N and ^{17}O chemical shift predictions in molecular crystals"

by J. Hartman, R. Kudla, G. Day, L. Mueller, and G. Beran

July 8, 2016

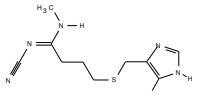
Contents

This supporting info provides 2-D structures for all species included in the ^{1}H, ^{15}N and ^{17}O benchmark sets. It also provides details regarding convergence with respect to the model parameters, additional tables of regression parameters, and error distributions from the test sets and the statistical cross validation sets. It also includes complete tables of the experimental and PBE0 chemical shifts used for each test set. CIF files containing the optimized geometries of all crystals in the benchmark tests sets are provided separately. Note that additional details regarding the composition and convergence properties of ^{13}C test set have been provided previously.[1]

1 Benchmark test set composition

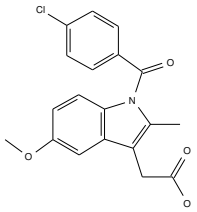
Figures S1–S3 depict the molecular structure of each crystal included in the ^{1}H, ^{15}N and ^{17}O benchmark sets. Reference codes are provided for structures obtained from the Cambridge Structure Database (CSD). If no reference code is provided, the crystal structure was obtained from the supporting information for the corresponding experimental NMR study. See Ref [1] for details of the ^{13}C test set.

Hydrogen test set



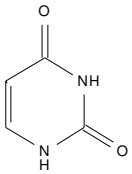
1. Cimetidine

CIMETD



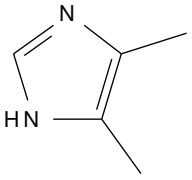
2. Indomethacin

INDMET

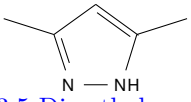


3. Uracil

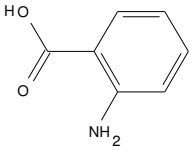
URACIL



4a. 4,5-dimethylimidazole

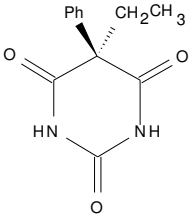


4b. 3,5-Dimethylpyrazole



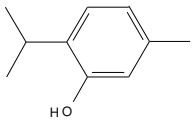
5. Anthranilic acid

AMBACO05



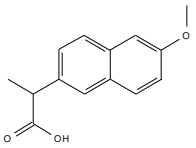
6. Phenobarbital

PHBARB06



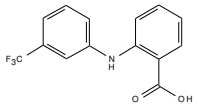
7. Thymol

IPMEPL



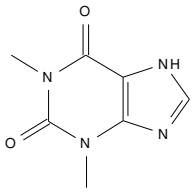
8. Naproxen

COYRUD11



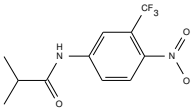
9. Flufenamic acid

FPAMCA11



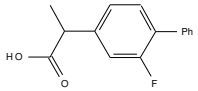
10. Theophylline

BAPLOT01



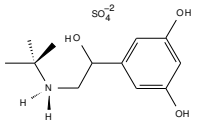
11. Flutamide

WEZCOT



12. Flurbiprofen

FLUBIP

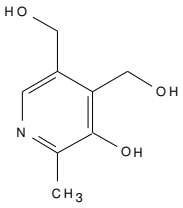


13. Terbutaline hemisulfate

ZIVKAQ

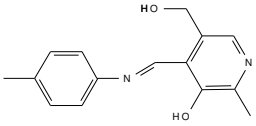
Figure S1: Crystal structures and corresponding CSD reference codes included in the ^{1}H benchmark set.

Nitrogen test set



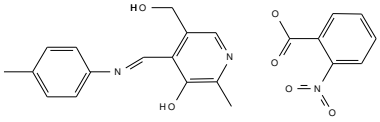
1. Pyridoxine

BITZAF



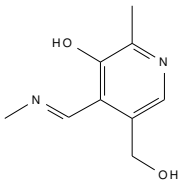
2. N-(Pyridoxylidene) tolylamine

GEHHAD



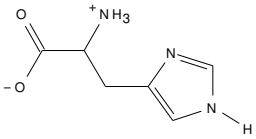
3. N-(Pyridoxylidenium)tolylamine 3.2-nitrobenzoate

GEHHEH



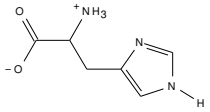
4. N-(Pyridoxylidene) methylamine

GEHHIL



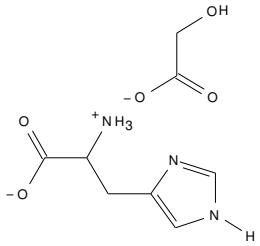
5. L-Histidine monoclinic

LHISTD02



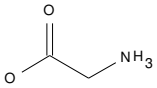
6.L-Histidine orthorombic

LHISTD13



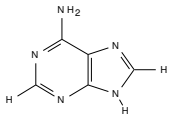
7. L-Histidine glycolate

TEJWAG



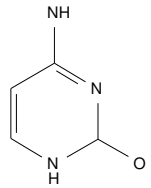
8. Glycine

GLYCIN03



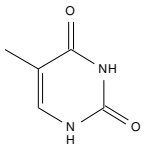
9. Adenine trihydrate

FUSVAQ01



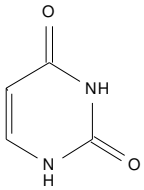
10. Cytosine

CYTSIN



11. Thymine

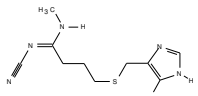
THYMIN01



12. Uracil

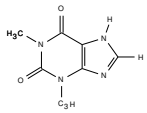
URACIL

Figure S2: Crystal structures and corresponding CSD reference codes included in the ^{15}N benchmark set.



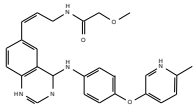
13. Cimetidine

CIMETD



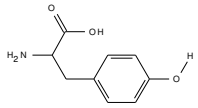
14. Theophylline

BAPLOT01



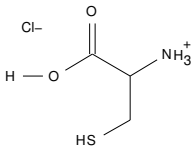
15. Compound **1**

f843



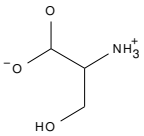
16. L-Tyrosine hydrochloride

LTYRHC10



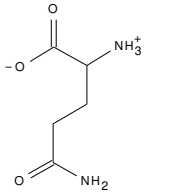
17. L-Cysteine hydrochloride monohydrate

CYSCLM11



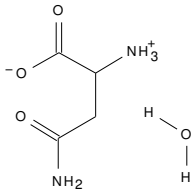
18. Serine

LSERIN01



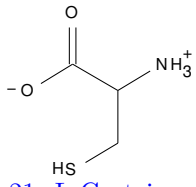
19. L-Glutamine

GLUTAM01



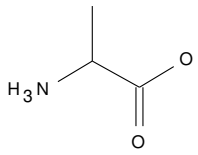
20. L-Asparagine monohydrate

ASPARM03



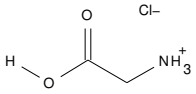
21. L-Cysteine

LCYSTN21



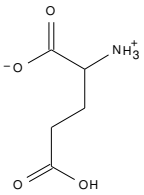
22. D-Alanine

ALUCAL04



23. Glycine hydrochloride

GLYHCL01

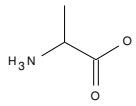


24. L-Glutamic acid beta-polymorph

LGLUAC11

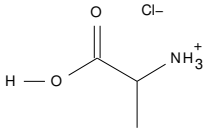
Figure S2: (cont.) Crystal structures and corresponding CSD reference codes included in the ^{15}N benchmark set.

Oxygen test set

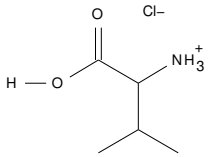


1. Alanine

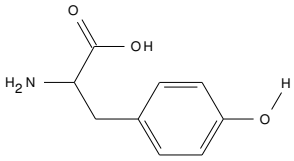
LALNIN12



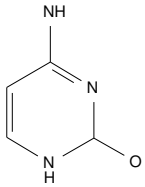
2. L-Alanine hydrochloride
ALAHCL



3. L-Valine hydrochloride
VALEHC11

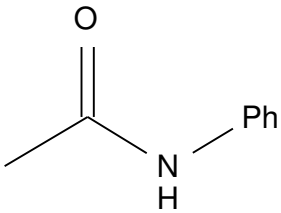


4. L-Tyrosine hydrochloride
LTYRHC10



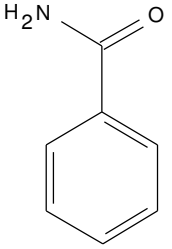
5. Cytosine

CYTSIN



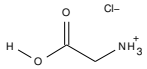
6. Acetanilide

ACANIL03

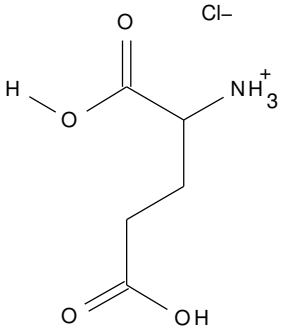


7. Benzamide

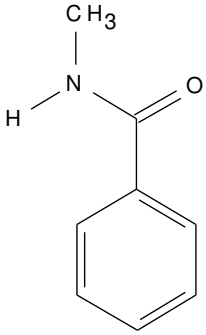
BZAMID07



8. Glycine hydrochloride
GLYHCL01

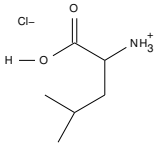


9. L-Glutamic acid hydrochloride
LGLUTA03

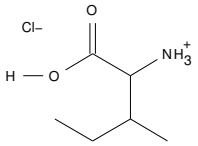


10. N-Methylbenzamide

MBNZAM10

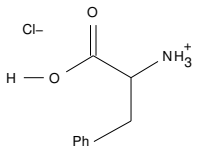


11. L-Leucine hydrochoride monohydrate
FEQYUW

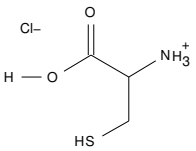


12. L-Isoleucine hydrochloride monohydrate
LILEUC10

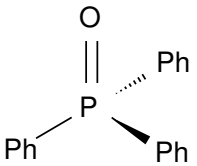
Figure S3: Crystal structures and corresponding CSD refcodes included in the ^{17}O benchmark set.



13. L-Phenylalanine hydrochloride
PHALNC01



14. L-Cysteine hydrochloride monohydrate
CYSCLM11



15. Triphenylphosphine oxide
TPEPHO02

Figure S3: (cont.) Crystal structures and corresponding CSD refcodes included in the ^{17}O benchmark set.

2 Analysis of isotropic shielding predictions

The following sections demonstrate convergence of the chemical shifts with respect to the basis set used to optimized the geometries, the locally dense basis set, and the 2-body cutoff distance. They also provide regression parameters and error distributions for the full test set and from the cross-validation experiments.

2.1 Effects of geometry optimization

Given the range of plane wave cutoff values reported in the literature and the sensitivity of the chemical shift to local geometry, an assessment of the plane wave cut-off values for all structures in the hydrogen, nitrogen and oxygen test-sets was performed. All-atom fixed-cell geometry optimizations were performed using Quantum Espresso as described in Section 3.2 of the main article using plane wave energy cutoffs of 40 Ry, 60 Ry and 80 Ry for all structures in each test set. Table S1 presents the root-mean-square (RMS) errors as well as linear regression parameters for each optimization scheme.

Table S1: RMS errors (in ppm) and linear regression parameters for the hydrogen, nitrogen and oxygen test sets optimized using plane wave cut offs of 40 Ry, 60 Ry and 80 Ry.

		40 Ry	**60 Ry**	**80 Ry**
Hydrogen	RMSE	0.49	0.33	0.34
	Slope	-0.9135	-0.9160	-0.9169
	Intercept	28.58	28.65	28.69
Nitrogen	RMSE	4.31	4.50	4.34
	Slope	-1.0296	-1.0340	-1.0311
	Intercept	197.80	198.55	198.09
Oxygen	RMSE	9.48	9.97	9.80
	Slope	-1.0568	-1.0607	-1.0607
	Intercept	269.48	270.19	270.18

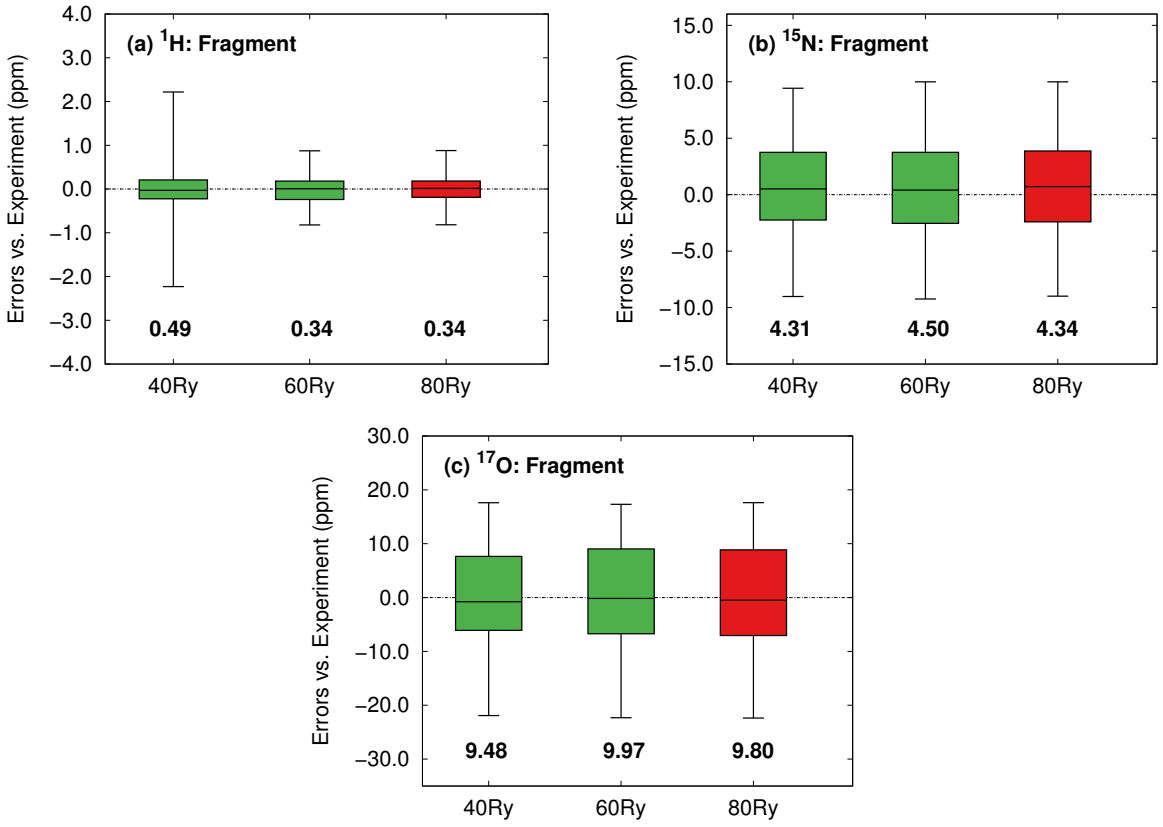


Figure S4: Error distributions relative to experiment for PBE0 mixed basis isotropic shielding predictions for the hydrogen, nitrogen and oxygen test sets optimized using plane wave cut offs of 40 Ry, 60 Ry and 80 Ry. The 80 Ry cut off used throughout this work is shown in red.

Although the linear regression parameters for ^{1}H are relatively insensitive to the basis set chosen for the geometry optimization, both the RMS errors (Table S1) and the error distributions shown in Figure S4a clearly illustrate the need for large plane wave cut offs of $\sim$ 60 Ry. On the other hand, ^{15}N and ^{17}O remain comparatively insensitive to minor changes in geometry. Both the overall RMS errors and linear regression parameters for ^{15}N and ^{17}O demonstrate minimal variation (Table S1). Further, the error distributions for ^{15}N and ^{17}O show a similar trend (Figures S4b and S4c).

Convergence with respect to **k**-point sampling was carried out for each structure included in this study. All-atom fixed-cell geometry optimizations were performed using an 80 Ry plane wave cut off and a variety of **k**-point spacing. Sufficiently dense **k**-point spacing was chosen for each of the optimized structures used in the ^{1}H, ^{13}C, ^{15}N and ^{17}O test sets, ensuring convergence to within $\sim$0.1 ppm. The 3x3x3 grid described in Section 3.2 of the main article proved more than sufficient for most structures. However, the following subset of structures required a denser **k**-point grid. The **k**-point grid used in the optimization is provided in parentheses following each structure: LHISTD02 (5×5×5), TEJWAG (5×4×3), GLUTAM01 (6×5×5), VALEHC11 (5×5×5), BZAMID07 (6×5×5), LGUTA03 (6×5×5).

2.2 Convergence with respect to basis set

Four different basis combinations were assessed to explore the effects of basis sets size on the resulting chemical shifts:

- Basis 1: A 6-31G* basis on all atoms within 4 Å of the asymmetric unit, and a 6-31G basis on the remaining atoms.
- Basis 2: A 6-311G** basis set is used on the asymmetric unit and all atoms within 2 Å, a 6-31G* basis is used on all atoms between 2 Å and 4 Å of the asymmetric unit, and a 6-31G basis is used on the remaining atoms.
- Basis 3: The locally dense basis described in Section 3.2 of the main article.
- Basis 4: A 6-311+G(2d,p) basis on all atoms.

Figure S5 plots the error distributions with respect to basis set for two-body fragment calculations using the PBE0 density functional and a 6 Å two-body distance cut off for each test set. The RMS errors are provided below each of the error distributions. Although significant variation in the absolute shieldings occur across basis sets, the linear regression largely compensates. Except for the smallest basis for ^{1}H, the overall RMS errors and error distributions vary only modestly with basis set size.

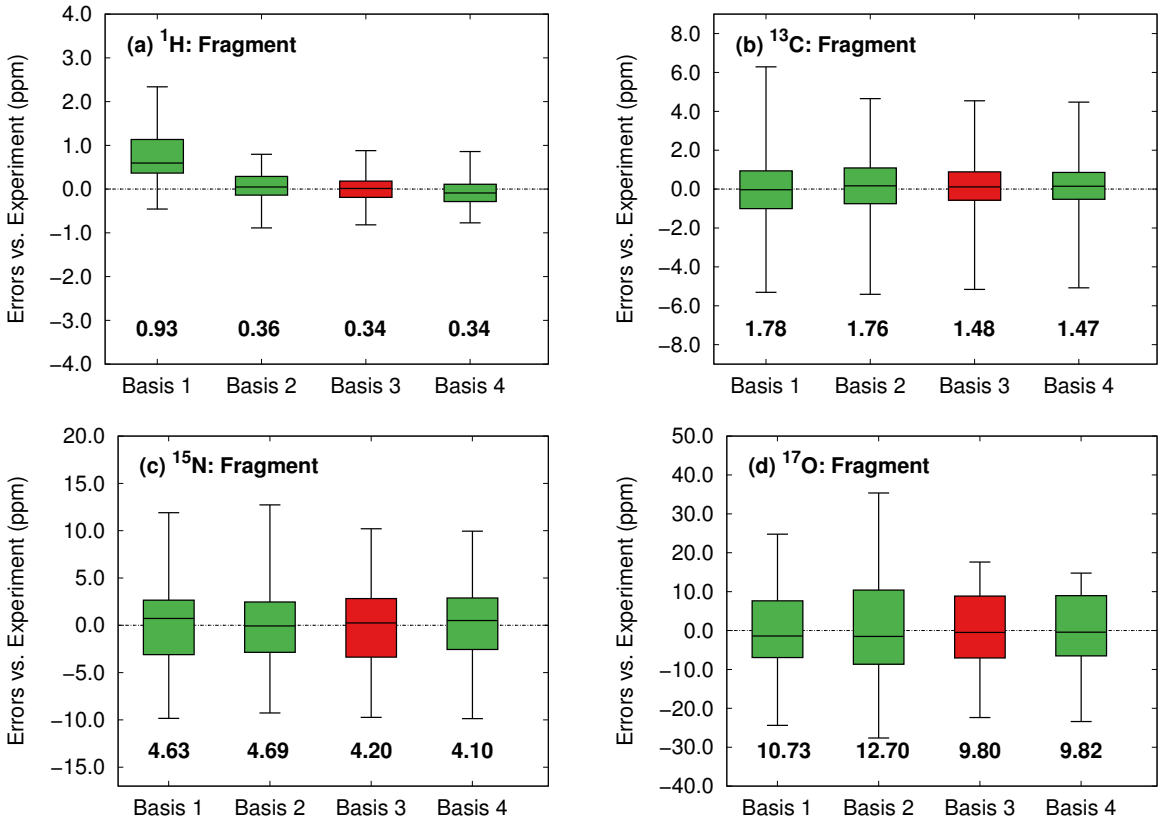


Figure S5: Error distributions relative to experiment for two-body fragment PBE0 isotropic shielding predictions for (a) ^{1}H, (b) ^{13}C, (c) ^{15}N and (d) ^{17}O. The results from the basis set used in the main article (Basis 3) are indicated in red.

2.3 Convergence with two-body cut off distance

Figure S6 plots the PBE0 error distributions with respect to the two-body cut off for the ^{1}H, ^{15}N, and ^{17}O test sets. All models with larger two-body cutoffs behave similarly. The resulting scaling parameters for each cut off are shown in in Table S2

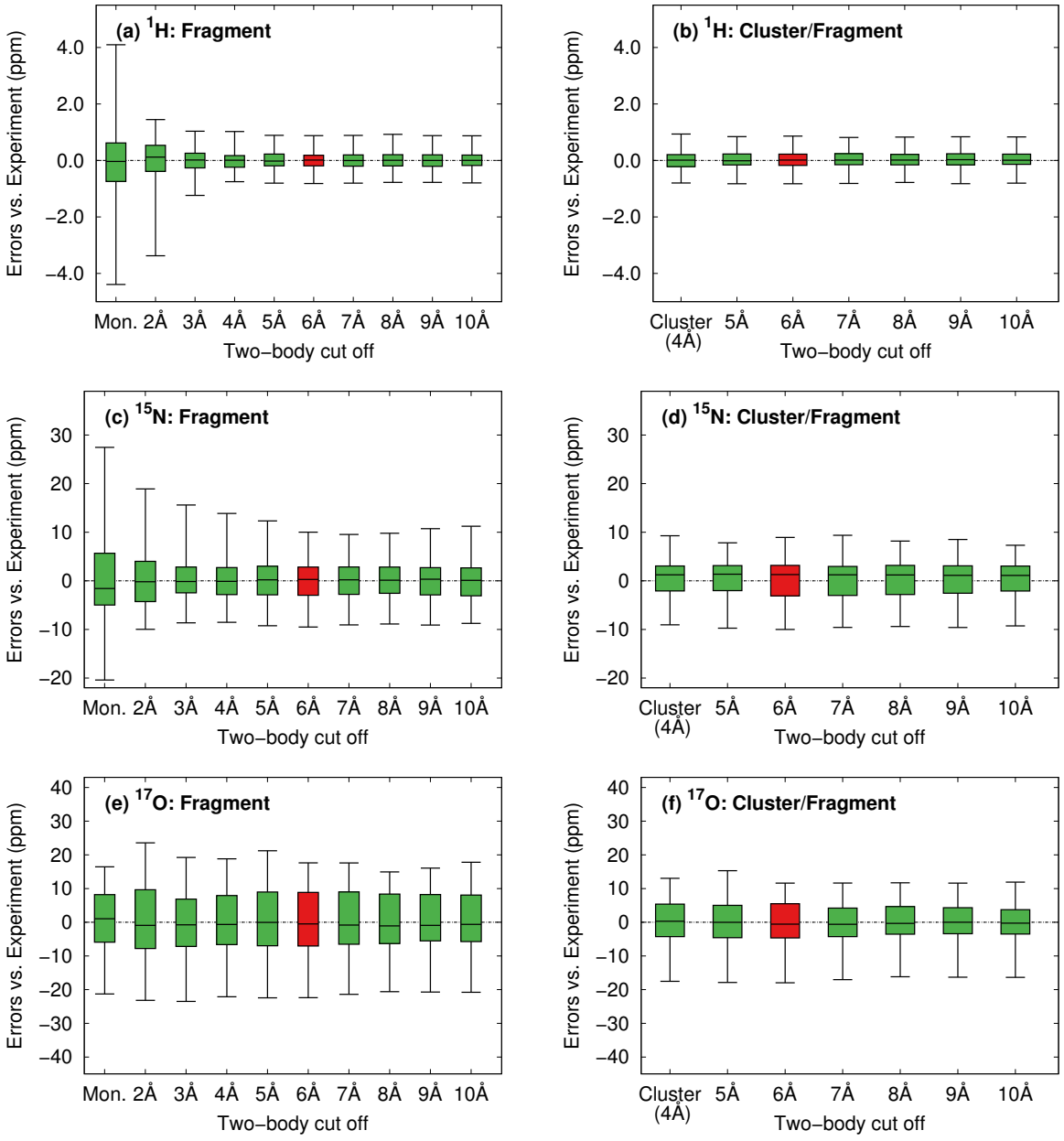


Figure S6: Error distributions relative to experiment for PBE0 mixed basis isotropic shielding predictions. The results from the 6 Å cut off used in the main article are indicated in red.

Table S2: Scaling parameters and RMS errors (in ppm) for PBE0 mixed basis calculations using different two-body distance cut offs.

		2-Body Fragment Model			Cluster/Fragment Model		
Atom	**2-body cutoff**	**RMSE**	**Slope**	**Intercept**	**RMSE**	**Slope**	**Intercept**
Hydrogen	Monomer	1.44	-1.0966	33.54			
	2 Å	0.82	-0.9000	27.99			
	3 Å	0.42	-0.9065	28.30			
	4 Å	0.36	-0.9198	28.66	0.37	-0.9141	28.46
	5 Å	0.33	-0.9177	28.67	0.33	-0.9121	28.47
	6 Å	0.34	-0.9169	28.69	0.35	-0.9111	28.49
	7 Å	0.33	-0.9191	28.75	0.34	-0.9134	28.55
	8 Å	0.33	-0.9199	28.76	0.34	-0.9142	28.55
	9 Å	0.34	-0.9202	28.76	0.34	-0.9144	28.56
	10 Å	0.33	-0.9211	28.80	0.34	-0.9153	28.60
Nitrogen	Monomer	8.39	-0.9198	196.71			
	2 Å	5.88	-0.9838	197.81			
	3 Å	4.41	-1.0182	197.97			
	4 Å	4.26	-1.0208	197.58	3.86	-1.0004	196.91
	5 Å	4.20	-1.0227	197.73	3.95	-1.0021	197.05
	6 Å	4.22	-1.0232	197.89	4.07	-1.0025	197.19
	7 Å	4.20	-1.0229	197.88	4.02	-1.0022	197.19
	8 Å	4.13	-1.0227	197.85	3.94	-1.0020	197.15
	9 Å	4.15	-1.0234	197.93	3.91	-1.0028	197.24
	10 Å	4.18	-1.0239	198.03	3.85	-1.0033	197.35
Oxygen	Monomer	9.63	-0.9008	279.48			
	2 Å	10.71	-0.9832	275.24			
	3 Å	10.22	-1.0447	271.19			
	4 Å	9.94	-1.0460	269.81	8.10	-1.0350	271.19
	5 Å	9.87	-1.0515	269.71	7.93	-1.0406	271.11
	6 Å	9.80	-1.0607	270.18	7.55	-1.0502	271.60
	7 Å	9.73	-1.0581	270.07	7.70	-1.0472	271.48
	8 Å	9.33	-1.0545	269.75	7.44	-1.0431	271.14
	9 Å	9.43	-1.0516	269.55	7.60	-1.0402	270.94
	10 Å	9.61	-1.0505	269.47	7.74	-1.0394	270.87

2.4 Cluster-only linear regression parameters

Although we do not generally recommend a cluster-only model, since modest improvements can be obtained at low computational cost by either a two-body fragment only model (^{1}H, ^{13}C, or ^{15}N) or by including longer-range two-body interactions via the combined cluster/fragment approach (^{17}O), Table S3 reports scaling parameters based on an electrostatically embedded 4 Å cluster model.

Table S3: Linear regression parameters for PBE0 mixed basis calculations using a 4 Å cluster model with electrostatic embedding.

Atom	**2-body cutoff**	**RMSE**	**Slope**	**Intercept**
Hydrogen	OPBE	0.39	-0.9351	29.02
	PBE	0.38	-0.9290	28.76
	TPSS	0.37	-0.9322	29.18
	TPSSh	0.37	-0.9270	29.05
	PBE0	0.37	-0.9141	28.46
	B3LYP	0.37	-0.9172	28.69
Carbon	OPBE	1.93	-1.0588	194.59
	PBE	2.14	-1.0236	180.18
	TPSS	2.22	-1.0415	185.68
	TPSSh	1.88	-1.0159	184.75
	PBE0	1.53	-0.9657	179.42
	B3LYP	1.54	-0.9682	173.63
Nitrogen	OPBE	6.44	-1.1041	215.04
	PBE	5.74	-1.0582	197.39
	TPSS	5.83	-1.0833	205.93
	TPSSh	5.22	-1.0595	205.23
	PBE0	3.85	-1.1041	196.87
	B3LYP	4.30	-0.9975	190.89
Oxygen	OPBE	9.11	-1.1205	281.17
	PBE	8.96	-1.1125	263.90
	TPSS	9.13	-1.1340	277.88
	TPSSh	8.71	-1.0994	279.57
	PBE0	8.10	-1.0350	271.19
	B3LYP	7.97	-1.0408	266.10

2.5 Test set error distributions

Figure S7 plots the error distributions for each test set. The error distributions are loosely normal, with ∼68% of the data falling within ±1 standard deviation and ∼95% of the data falling within ±2 standard deviations of the mean.

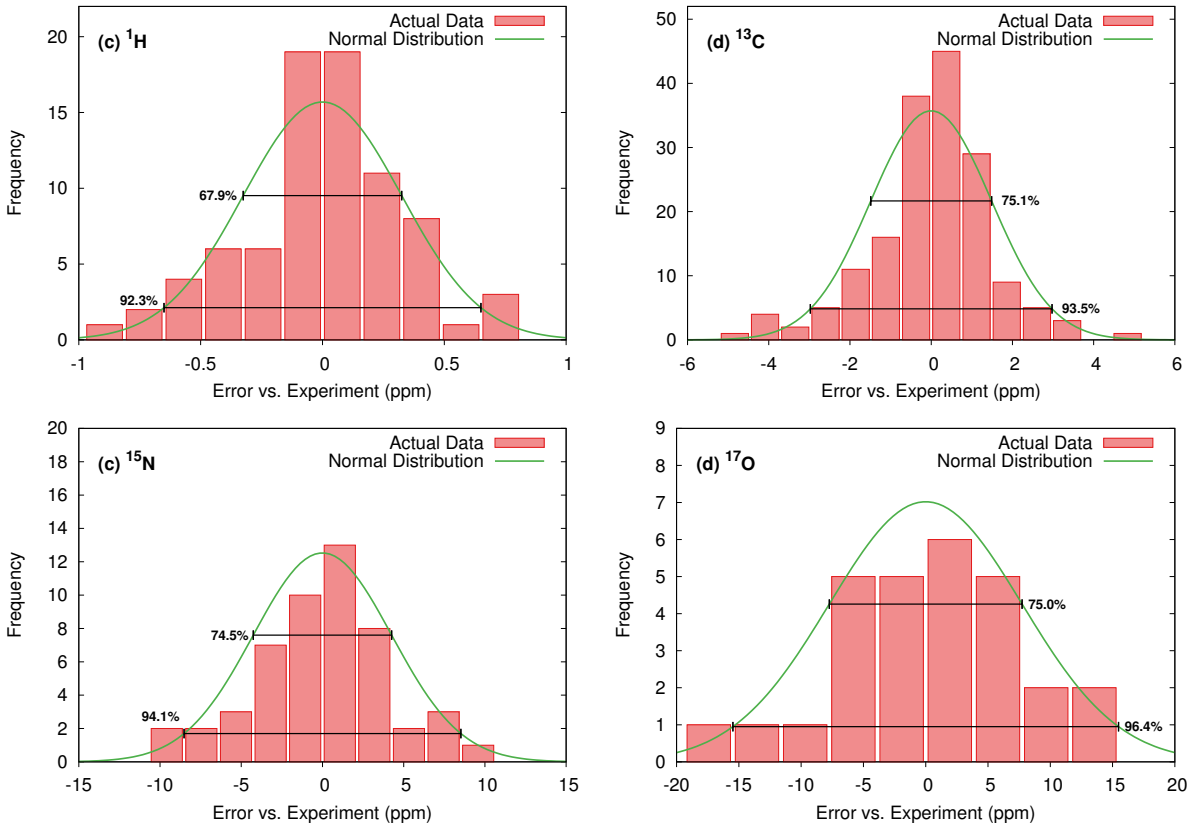


Figure S7: Error distributions relative to experiment for PBE0 mixed basis isotropic shielding predictions using the fragment 2-body (^{1}H, ^{13}C, or ^{15}N) or cluster/fragment (^{17}O) models. Histogram bin widths correspond to half a standard deviation, and fraction of the data falling within ±1 or ±2 standard deviations is noted.

2.6 Statistical cross-validation error distributions

Figure S8 plots the complete error distributions over all possible N-choose-5 cross-validation experiments for each test set. These errors correspond to the errors for the predictions of the five crystals not included in the fit for each possible composition of the training and validation subsets. The hydrogen, carbon, and nitrogen results are based on the 2-body fragment model, while the oxygen results are for the combined cluster/fragment model. The widths of these distributions are fairly similar to those of the error distributions obtained by simply fitting to the entire set (Figure S7).

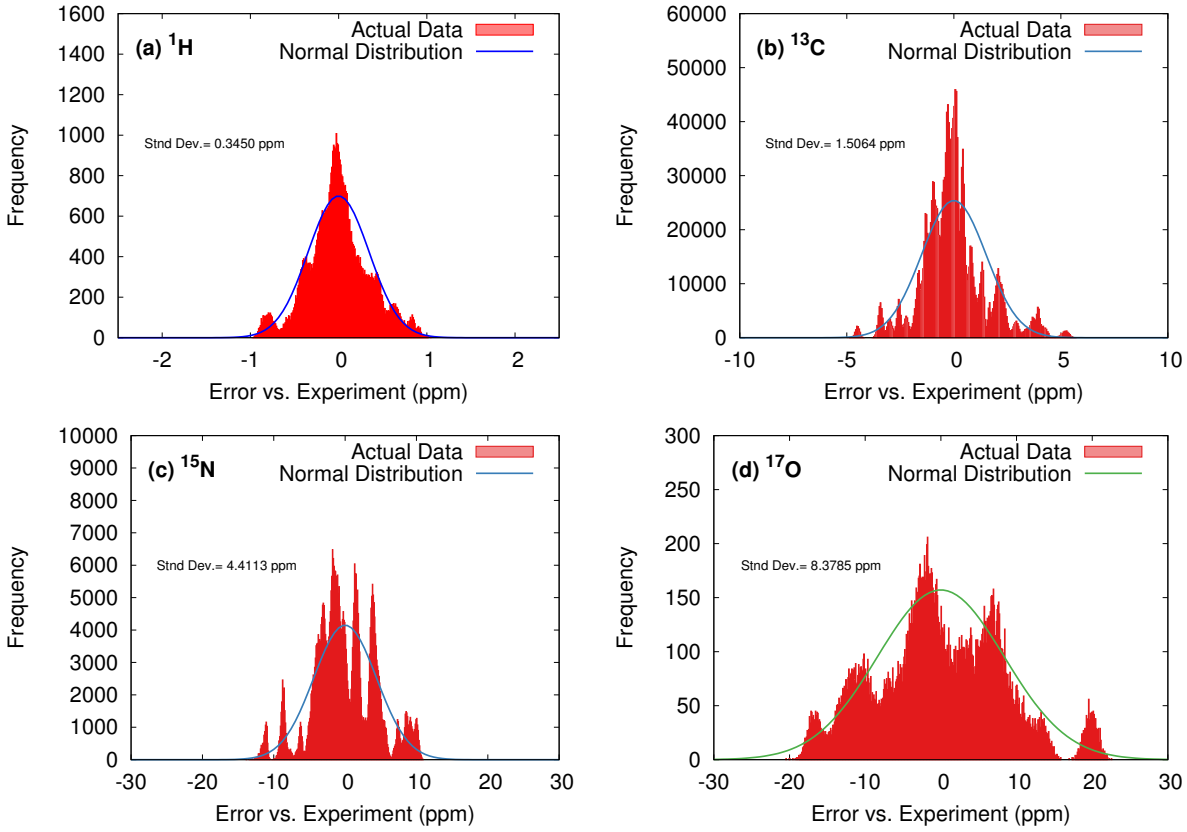


Figure S8: Individual errors in the predicted isotropic shifts relative to experiment for all N-choose-5 permutations for each of the respective test sets.

2.7 Convergence of GIPAW chemical shifts

The convergence of isotropic shieldings calculated using the GIPAW approach was examined for three crystal structures selected from the benchmark sets. Table S4 shows convergence with respect to the planewave basis set cutoff energy. Absolute chemical shieldings of ^{13}C, ^{15}N and ^{17}O have converged to within 0.06 ppm once the basis set is increased to 1050 eV and 0.01 ppm using a 1150 eV cutoff, while convergence of ^{1}H is achieved at a smaller basis set (0.04 ppm convergence using a 850 eV cutoff, 0.01 ppm using a 950 eV cutoff). Comparison of relative shieldings shows that their values are converged earlier than absolute shieldings; 0.05 ppm convergence is achieved for all nuclei using an 850 eV basis set cutoff, which we have chosen for all benchmark calculations. Convergence of calculated shieldings with respect to the density of the **k**-point grid is shown in Table S5, which demonstrates only weak dependence of results on the chosen **k**-point grid. A 0.05

Å^{-1} target spacing between **k**-points reproduces the shieldings calculated using a much denser 0.03 Å^{-1} grid to within 0.03 ppm.

Table S4: Calculated chemical shielding of selected nuclei in the crystal structures of L-alanine (LALNIN12), benzamide (BZAMID07) and flutamide (WEZCOT), showing convergence with respect to the planewave basis set cutoff energy. Atoms are labeled as in the reported crystal structures. Errors are calculated with respect to the largest basis set in each case (1250 eV for O, N, C, 1050 eV for H). All calculations were performed with a **k**-point grid chosen to have a maximum spacing of 0.04 Å^{-1}.

		Basis Set Cutoff (in eV)							
Structure	**Atom**	**550**	**650**	**750**	**850**	**950**	**1050**	**1150**	**1250**
BZAMID07	O				calculated shielding (in ppm)				
		-36.41	41.45	-42.36	-42.23	-42.02	-41.89	-41.84	-41.83
					error with respect to 1250 eV				
		5.41	0.38	-0.53	-0.40	-0.19	-0.06	-0.01	-
LALNIN12	O1				calculated shielding (in ppm)				
		-40.26	-45.23	-46.15	-46.01	-45.79	-45.67	-45.62	-45.61
					error with respect to 1250 eV				
		5.35	0.38	-0.54	-0.40	-0.18	-0.06	-0.01	-
LALNIN12	O2				calculated shielding (in ppm)				
		-20.88	-25.64	-26.52	-26.41	-26.19	-26.07	-26.02	-26.02
					error with respect to 1250 eV				
		5.14	0.38	-0.50	-0.39	-0.17	-0.05	0.00	-
BZAMID07	N				calculated shielding (in ppm)				
		112.78	110.05	109.53	109.58	109.70	109.77	109.80	109.81
					error with respect to 1250 eV				
		2.97	0.24	-0.28	-0.23	-0.11	-0.04	-0.01	-
LALNIN12	N				calculated shielding (in ppm)				
		186.34	184.17	183.74	183.78	183.87	183.93	183.95	183.96
					error with respect to 1250 eV				
		2.38	0.21	-0.22	-0.18	-0.09	-0.03	-0.01	-
BZAMID07	C1 (aromatic)				calculated shielding (in ppm)				
		33.29	32.73	32.70	32.74	32.76	32.76	32.75	32.74
					error with respect to 1250 eV				
		0.55	-0.01	-0.04	0.00	0.02	0.02	0.01	-
BZAMID07	C4 (amide)				calculated shielding (in ppm)				
		-2.07	-2.77	-2.83	-2.79	-2.78	-2.79	-2.80	-2.81
					error with respect to 1250 eV				
		0.74	0.04	-0.02	0.02	0.03	0.02	0.01	-
LALNIN12	C1 (carboxyl)				calculated shielding (in ppm)				

Continued on next page

Table S4 – *Continued from previous page*

Structure	Atom	Basis Set Cutoff (in eV) 550	650	750	850	950	1050	1150	1250
		-9.52	-10.16	-10.26	-10.21	-10.20	-10.22	-10.23	-10.24
					error with respect to 1250 eV				
		0.72	0.08	-0.02	0.03	0.04	0.02	0.01	-
WEZCOT	H1 (aromatic)				calculated shielding (in ppm)				
		22.96	23.07	23.14	23.18	23.21	23.22		
					error with respect to 1050 eV				
		-0.26	-0.15	-0.08	-0.04	-0.01	-		
WEZCOT	H4 (N-H)				calculated shielding (in ppm)				
		21.08	21.18	21.24	21.28	21.30	21.31		
					error with respect to 1050 eV				
		-0.23	-0.13	-0.07	-0.03	-0.01	-		
WEZCOT	H5 (C-H)				calculated shielding (in ppm)				
		28.25	28.37	28.44	28.48	28.51	28.52		
					error with respect to 1050 eV				
		-0.27	-0.15	-0.08	-0.04	-0.01	-		

Table S5: Calculated chemical shielding (in ppm) of selected nuclei in the crystal structures of L-alanine (LALNIN12) and benzamide (BZAMID07) showing convergence with respect to the density of the electronic **k**-point grid. All calculations were performed with a basis set cutoff of 1050 eV. For BZAMID07, the chosen grids correspond to 4, 4, 15 and 24 **k**-points in the irreducible Brillouin Zone for 0.06, 0.05, 0.04 and 0.03 $Å^{-1}$, respectively. For LALNIN12, the grids correspond to 4, 4, 18 and 18 **k**-points in the irreducible Brillouin Zone.

		Target spacing between k-points ($Å^{-1}$)			
Structure	**Atom**	**0.06**	**0.05**	**0.04**	**0.03**
BZAMID07	O	-41.92	-41.88	-41.89	-41.88
LALNIN12	O1	-45.71	-45.70	-45.67	-45.68
LALNIN12	O2	-26.09	-26.11	-26.07	-26.09
BZAMID07	N	109.70	109.75	109.77	109.76
LALNIN12	N	183.88	183.89	183.93	183.91
BZAMID07	C1 (aromatic)	32.72	32.73	32.76	32.76
BZAMID07	C4 (amide)	-2.79	-2.81	-2.79	-2.79
LALNIN12	C1 (carboxyl)	-10.21	-10.25	-10.22	-10.23

3 Tabulated experimental and predicted chemical shifts

Tables S4–S7 list the experimental and PBE0-predicted chemical shifts for every species included in the ^{1}H, ^{13}C, ^{15}N, and ^{17}O test sets, respectively. Note, due to a slightly different geometry optimization procedure, the ^{13}C chemical shifts in Table S5 differ slightly from those published previously.[1]

Table S6: Comparison of experimental and predicted isotropic **^{1}H chemical shifts** for the 2-body, cluster, and combined cluster/2-body models with charge embedding using the PBE0 density functional and those obtained using GIPAW with the PBE density functional (in ppm). The raw chemical shieldings can be obtained from the empirically scaled chemical shieldings reported here using the linear regression parameters reported in this work. All shieldings are reported relative to TMS with adamantane at 1.87 ppm.

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
CIMETD	2.24	1.83	1.70	1.80	2.17
	2.24	1.94	1.85	1.92	2.20
	7.64	7.07	7.10	7.18	7.42
	8.44	8.42	8.38	8.38	8.46
	9.94	10.07	9.79	9.91	9.74
	11.84	11.54	11.41	11.52	12.27
INDMET	5.8	5.72	5.70	5.64	5.74
	6.1	6.05	6.07	6.05	5.96
	5.8	5.78	5.79	5.75	5.72
	1.7	1.69	1.74	1.67	1.89
	2.2	2.18	2.25	2.17	2.58
	1.8	1.87	2.04	1.79	2.12
	7.3	6.95	7.02	6.89	6.90
	5.7	5.73	5.95	5.75	5.53
	7.2	7.09	7.04	6.93	6.81
	7.3	7.44	7.44	7.30	6.97
URACIL	11.2	11.48	11.29	11.40	12.00
	10.8	10.98	10.92	11.06	11.04
	7.5	7.88	7.79	7.93	7.58
	6	6.41	6.38	6.50	6.06
4,5-Dimethylimidazole	4.8	4.57	4.41	4.61	4.56
	1.4	1.39	1.24	1.38	1.47
	0.7	1.51	1.39	1.51	1.53
	13	13.18	13.23	13.41	13.46
3,5-Dimethylpyrazole	5.2	5.54	5.51	5.51	5.60
	1.5	1.88	1.82	1.91	2.05
	1.4	2.15	2.14	2.16	2.05
	15	14.69	14.71	14.75	14.54
AMBACO05	5.8	5.87	6.00	5.83	5.58
	6.8	7.03	7.12	7.00	6.72
	5.4	5.72	5.97	5.79	5.91
PHBARB06	10.3	10.46	10.43	10.52	11.45

Continued on next page

Table S6 – *Continued from previous page*

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
	8.1	8.07	7.98	8.06	8.97
	10.3	10.35	10.25	10.31	11.13
	6.9	6.93	6.90	6.95	6.73
	2.7	1.99	1.95	2.02	2.17
	1.7	1.71	1.66	1.71	1.81
	0.6	-0.17	-0.24	-0.13	0.29
IPMEPL	0.42	1.07	1.09	1.10	1.29
	1.05	1.56	1.41	1.57	1.63
	1.45	0.57	0.42	0.57	0.71
	3.38	3.37	3.28	3.31	3.35
	5.4	6.11	5.91	6.07	5.76
	9.99	10.24	10.57	10.58	10.36
COYRUD11	7	7.56	7.62	7.55	7.15
	6.1	6.45	6.73	6.45	5.93
	3.8	3.67	3.92	3.62	3.40
	4.5	4.79	4.81	4.74	4.55
	4.1	4.45	4.72	4.49	4.20
	5.9	5.83	6.06	5.85	5.34
	3.2	3.78	3.94	3.81	3.76
	1.8	2.01	2.08	2.09	2.13
	2.3	2.24	2.34	2.31	2.55
FPAMCA11	8.3	8.13	7.90	8.07	7.66
	6	6.06	5.94	6.07	5.73
	5.4	5.32	5.39	5.39	4.99
	6.8	6.91	6.90	6.92	6.79
	9.6	9.59	9.42	9.52	9.69
	6.9	6.85	6.72	6.79	6.35
	6.2	6.41	6.31	6.37	6.00
	5.9	6.04	5.94	5.95	5.59
	7.3	7.33	7.29	7.32	6.90
BAPLOT01	14.6	14.55	14.83	14.52	14.53
	7.7	7.22	7.56	7.22	7.15
	3.4	3.34	3.47	3.37	3.45
WEZCOT	7.1	6.71	6.77	6.70	6.95
	9.9	9.51	9.43	9.48	9.04
	8	7.82	7.91	7.88	7.34
	8	7.95	7.94	7.90	8.62
	2	1.80	1.88	1.83	2.33
	1.2	1.05	1.13	1.16	1.34
FLUBIP	0.9	0.94	0.96	0.94	1.00
	2.9	2.67	2.72	2.66	2.51
	6.7	6.57	6.51	6.50	6.16
ZIVKAQ	1.33	1.01	1.08	1.13	1.32
	4.73	4.39	4.37	4.43	4.64

Continued on next page

Table S6 – *Continued from previous page*

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
	6.83	6.67	6.49	6.57	6.37
	6.83	6.61	6.53	6.60	6.36
	6.83	6.71	6.60	6.70	6.61
	10.93	10.15	10.00	10.08	10.32

Table S7: Comparison of experimental and predicted isotropic ^{13}**C chemical shifts** for the 2-body, cluster, and combined cluster/2-body models with charge embedding using the PBE0 density functional and those obtained using GIPAW with the PBE density functional (in ppm). The raw chemical shieldings can be obtained from the empirically scaled chemical shieldings reported here using the linear regression parameters reported in this work. All shieldings are reported relative to TMS.

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
GLYCIN03	176.2	179.48	178.21	178.90	177.58
	43.5	43.43	43.65	43.80	40.71
LALNIN12	176.8	180.96	180.06	180.24	179.33
	50.9	51.23	51.15	51.15	49.09
	19.8	20.80	20.56	20.59	16.86
LSERIN01	175.1	177.96	178.32	177.78	176.28
	55.6	55.57	55.78	55.67	53.99
	62.9	64.26	63.98	63.94	65.39
LTYROS11	176	178.82	178.10	178.12	176.57
	130.3	129.91	128.33	129.24	129.75
	117.2	116.12	114.84	115.41	115.54
	54.7	56.71	56.82	56.90	55.61
	131	131.41	132.42	131.94	130.04
	117.2	119.38	120.75	119.56	117.13
	155.7	155.56	155.94	155.83	157.39
	35.8	38.06	37.79	37.88	36.18
	123	122.11	120.98	121.69	121.38
LCYSTN21	175.1	175.01	175.21	175.23	173.47
	53.7	54.50	54.87	54.77	53.56
	35.4	31.94	31.83	31.81	29.69
MGLUCP11	101	100.31	100.34	100.46	105.46
	72.3	70.95	70.34	70.59	72.33
	74.6	73.99	74.10	73.88	75.86
	72.5	71.37	71.78	71.79	73.62
	75.3	73.69	73.61	73.76	75.74
	63.8	63.78	63.01	63.15	63.99
	56.5	57.14	56.96	56.82	57.04
MBDGAL02	105.7	105.03	104.79	104.86	110.34

Continued on next page

Table S7 – *Continued from previous page*

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
	71.2	70.72	70.58	70.66	72.39
	72.1	72.24	72.22	72.35	74.33
	69.3	70.50	70.26	70.17	71.92
	75.6	74.43	74.43	74.51	76.67
	62.8	62.46	62.88	62.89	63.67
	57.6	58.93	59.10	59.18	59.04
MEMANP11	99.6	99.62	99.80	99.86	104.69
	71.3	70.95	70.78	70.84	72.83
	71.7	71.82	72.09	72.12	73.69
	64.8	64.48	64.67	64.77	65.60
	71.9	72.33	72.30	72.35	73.89
	58.9	58.46	58.57	58.52	58.76
	54.9	56.51	56.23	56.32	55.31
MGALPY01	100.4	99.88	99.93	99.91	105.34
	67.6	67.27	67.13	67.24	68.28
	72.6	73.04	73.58	73.46	74.71
	70	69.66	69.99	69.87	71.64
	72.9	72.01	72.41	72.43	74.02
	61.4	60.87	61.10	61.00	62.31
	55.2	56.50	56.16	56.12	55.23
XYLOBM01	104.2	103.30	103.09	103.27	108.59
	72.2	71.14	71.58	71.48	73.07
	78.2	76.89	76.61	76.76	79.17
	69.5	69.60	69.07	69.10	70.72
	66.9	66.68	66.40	66.44	67.88
	57.3	57.72	57.48	57.54	56.89
FRUCTO02	65.4	65.05	65.44	65.43	66.13
	99.7	101.54	101.44	101.54	106.54
	67.2	67.35	67.33	67.40	68.37
	69	67.78	67.52	67.68	68.98
	71.4	72.17	71.84	72.00	74.07
	64.9	65.78	65.63	65.57	66.66
RHAMAH12	94.5	94.59	94.80	94.77	99.37
	72.2	71.98	72.14	72.20	73.82
	71	69.67	69.86	69.86	71.46
	72.5	72.91	72.91	73.05	74.34
	69.8	70.19	70.07	70.18	71.80
	17.8	18.29	18.44	18.43	13.51
SUCROS04	93.3	92.80	92.95	92.93	96.57
	66	64.77	64.67	64.67	65.29
	73.7	72.98	73.14	73.15	75.06
	102.4	103.25	103.39	103.43	107.70
	72.8	71.91	72.04	72.04	73.99
	82.9	82.00	81.95	81.97	85.09
	67.9	67.20	67.48	67.57	68.87
	71.8	70.83	70.68	70.64	71.85

Continued on next page

Table S7 – *Continued from previous page*

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
	73.6	73.42	73.67	73.54	75.30
	81.8	80.39	80.53	80.48	82.73
	60	58.82	58.98	58.97	59.57
	61	61.23	61.22	61.19	61.67
GLUTAM01	177	175.85	176.20	175.98	174.26
	54	54.11	54.65	54.34	52.65
	26	26.41	26.64	26.62	23.99
	29	30.84	30.28	30.37	26.60
	174	175.98	175.43	175.92	174.71
ASPARM03	176.4	177.55	178.62	178.26	177.49
	51.8	50.88	51.26	51.09	49.12
	36.1	36.20	35.55	35.55	32.93
	177.1	174.18	174.96	174.87	173.22
LSERMH10	175.6	177.64	177.69	178.31	177.04
	58.3	58.48	58.26	58.37	57.63
	61.8	61.52	61.70	61.63	62.58
LTHREO01	170.0	173.95	174.79	174.74	172.18
	60.2	59.98	60.62	60.47	59.45
	65.4	67.03	67.20	67.32	68.95
	18.9	21.27	21.21	21.28	16.84
NAPHTA36	126	126.02	125.76	125.88	124.60
	129.3	130.12	129.67	129.93	128.97
	134.9	133.33	133.08	133.31	132.64
	129.9	130.32	130.01	130.17	129.32
	125.4	125.32	125.00	125.13	124.07
ACENAP03	148.1	149.17	149.65	149.30	150.32
	120.3	120.09	120.12	119.90	119.06
	129.4	130.11	130.22	129.94	129.21
	122.3	122.46	122.61	122.43	121.37
	131.9	130.19	130.44	130.11	129.67
	139.9	138.16	138.20	137.86	138.18
	29.5	30.78	30.80	30.77	27.11
TRIPHE11	126.4	126.25	126.38	126.37	125.45
	129.5	128.54	128.45	128.58	128.12
	124.5	124.84	125.06	125.25	123.61
	125.9	125.60	125.58	125.76	124.52
	127.5	127.05	127.10	127.25	126.23
	122.3	122.33	122.06	122.19	120.93
	130.2	129.43	129.27	129.41	129.06
	129.5	128.22	128.11	128.19	128.09
	120.9	120.48	120.35	120.42	119.16
	125.9	125.92	125.60	125.61	124.91
	121.7	121.82	122.19	122.20	120.79
	129.5	128.61	128.41	128.46	128.29

Continued on next page

Table S7 – *Continued from previous page*

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
	129.5	128.29	128.39	128.41	128.01
	122.3	122.94	122.75	122.66	121.27
	126.9	126.58	126.69	126.45	125.53
	126.9	128.33	128.40	128.21	126.90
	123.8	123.66	124.06	124.06	122.63
	129.8	129.06	129.13	129.19	128.79
HXACAN09	152.3	150.22	150.69	150.30	152.09
	116.4	115.44	117.01	116.22	115.01
	120.6	120.76	121.36	120.76	119.21
	133.1	130.90	130.93	131.26	131.21
	123.4	123.97	123.09	123.62	122.71
	115.7	116.19	114.50	114.84	113.95
	169.8	169.68	169.62	169.74	166.80
	23.8	27.34	27.54	27.25	22.16
INDMET	167.7	169.64	169.51	169.35	168.34
	136.7	132.16	132.29	131.99	132.88
	134.4	132.81	132.74	132.92	132.35
	129.2	130.17	130.28	130.33	130.41
	140.3	144.05	144.48	144.41	145.06
	127	126.50	127.11	126.76	126.47
	132	132.78	133.00	132.52	132.60
	13.7	16.12	16.53	16.11	11.38
	28.2	30.38	30.29	30.14	26.41
	179.2	184.36	184.25	183.94	183.48
	138	137.75	138.70	138.07	137.86
	55.2	55.50	55.19	55.12	54.39
	112.6	112.48	113.37	112.84	113.61
	131.1	129.72	130.31	129.92	130.32
	97.9	97.37	97.45	97.01	94.52
	156.7	155.65	155.42	155.41	157.33
	112.6	111.85	110.97	111.12	109.64
	115.7	116.03	115.87	116.18	114.26
	131.1	129.19	128.99	128.98	129.06
SULAMD06	127.1	123.81	124.44	124.16	125.63
	129.5	129.67	129.86	129.89	128.05
	117.1	114.58	114.21	114.51	114.16
	153.4	150.94	150.37	150.68	149.22
	112.3	110.71	110.21	110.31	110.12
	129.5	131.61	131.21	131.05	130.31
ADENOS12	154.8	155.07	155.01	155.11	152.39
	148.5	146.00	146.03	146.21	145.48
	119.7	119.19	119.35	119.50	120.22
	155.2	152.28	152.25	152.39	150.34
	137.8	137.10	136.28	136.62	134.92
	92.3	92.19	92.29	92.44	95.93
	71.2	75.15	75.13	75.31	77.06
	75	71.57	71.59	71.72	73.22
	84.9	84.82	84.72	84.84	87.19

Continued on next page

Table S7 – *Continued from previous page*

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
	62.7	62.79	62.59	62.70	63.11
PERYTO10	50.2	50.59	50.64	50.65	50.20
	58.4	58.42	58.44	58.45	58.97

Table S8: Comparison of experimental and predicted isotropic ^{15}**N chemical shifts** for the 2-body, cluster, and combined cluster/2-body models with charge embedding using the PBE0 density functional and those obtained using GIPAW with the PBE density functional (in ppm). The raw chemical shieldings can be obtained from the empirically scaled chemical shieldings reported here using the linear regression parameters reported in this work. All shieldings are reported relative to NH_4Cl with $NH_3(l)$ at 39.3 ppm.

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
BITZAF	249.5	247.58	245.73	245.30	241.16
GEHHAD	261.8	265.25	266.93	265.54	262.80
	253.6	247.25	244.16	244.50	245.17
GEHHEH	187.4	179.30	186.15	184.10	176.34
	261.0	250.79	260.32	264.07	249.55
GEHHIL	268.5	269.91	268.83	267.69	269.81
	261.2	261.76	258.40	259.02	256.76
LHISTD02	132.6	141.54	139.68	141.38	126.83
	210.8	209.71	209.06	207.41	211.79
LHISTD13	132.4	139.70	136.98	138.56	139.09
	210.6	211.77	212.34	210.70	212.39
TEJWAG	143.9	145.09	145.40	145.25	147.80
GLYCIN03	-6.5	-7.20	-9.49	-9.82	-12.77
FUSVAQ01	183.2	181.82	183.14	180.70	183.20
	174.2	173.91	171.55	173.10	172.97
	192.2	191.70	188.09	189.53	190.78
	120.2	122.87	126.34	127.08	127.46
	50.2	46.21	49.15	49.53	50.94
CYTSIN	110.2	114.88	112.79	114.42	117.21
	165.2	160.43	161.61	160.33	165.91
	54.2	57.53	55.45	53.81	52.86
THYMIN01	119.2	120.53	121.42	121.22	123.70

Continued on next page

Table S8 – *Continued from previous page*

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
	90.2	93.89	95.48	95.15	96.98
URACIL	96.2	105.90	105.43	106.38	108.56
	120.2	121.71	121.71	121.71	124.66
CIMETD	130.5	136.00	134.75	135.93	136.80
	213.1	214.03	214.39	213.35	212.86
	56.6	58.48	60.82	60.00	59.75
	43.5	43.70	43.74	44.85	43.22
	45.8	37.26	41.24	41.21	41.85
	149.9	146.29	149.05	148.38	146.27
BAPLOT01	114.7	122.84	123.72	123.51	126.14
	72.7	76.18	78.26	78.67	80.32
	122.7	126.27	127.10	126.36	127.51
	178.7	177.37	176.46	176.91	178.93
Compound 2	76.4	73.94	75.35	75.35	79.79
	81.5	78.64	79.21	79.58	78.31
	202.5	199.57	198.60	198.44	202.65
	223.4	219.66	216.96	217.53	222.82
	285.1	285.23	283.38	283.50	284.70
LTYRHC10	8.0	5.82	5.18	5.20	3.74
CYSCLM11	1.5	2.34	0.22	0.19	-2.29
LSERIN01	-4.1	-6.14	-6.68	-6.57	-9.57
GLUTAM01	-1.3	-2.71	-3.08	-3.37	-6.16
	72.1	75.82	73.26	73.24	75.15
ASPARM03	0.7	-0.99	-2.43	-2.54	-4.99
	74.9	79.42	75.11	75.41	77.66
LCYSTN21	-0.4	-4.96	-4.35	-4.60	-7.46
ALUCAL04	3.0	-0.43	0.41	0.31	-1.86
GLYHCL01	-1.7	-4.82	-3.64	-3.81	-6.30
LGLUAC11	-0.4	-1.76	-2.81	-2.80	-5.36

Table S9: Comparison of experimental and predicted isotropic ^{17}**O chemical shifts** for the 2-body, cluster, and combined cluster/2-body models with charge embedding using the PBE0 density functional and those obtained using GIPAW with the PBE density functional (in ppm). The raw chemical shieldings can be obtained from the empirically scaled chemical shieldings reported here using the linear regression parameters reported in this work. All shieldings are reported relative to liquid H_2O.

Crystal	Exp. Shifts Isotropic (ppm)	PBE0 2bd 6 Å	PBE0 Cluster 4 Å	PBE0 Cluster/ Fragment	PBE GIPAW
LALNIN12	285.0	299.80	298.99	297.42	297.39
	268.0	275.98	273.00	274.40	276.50
ALAHCL	327.8	328.43	328.31	329.90	325.00
	176.7	182.44	185.40	185.53	185.69
VALEHC11	351.0	338.35	349.52	347.82	354.01
	181.0	203.38	199.32	199.50	190.23
LTYRHC10	327.0	331.77	332.04	330.01	328.45
	183.0	182.34	183.13	183.60	184.71
	83.0	74.71	78.46	76.83	77.06
CYTSIN	230.0	212.38	217.55	218.81	221.28
ACANIL03	330.0	322.65	319.63	323.39	320.01
BZAMID07	300.0	288.76	289.72	290.58	293.32
GLYHCL01	336.0	344.41	332.80	335.78	329.94
	185.0	175.56	183.24	181.98	183.52
LGLUTA03	172.5	177.64	179.51	178.85	176.09
	322.0	308.08	307.37	310.82	309.89
	315.0	302.36	298.60	300.14	299.29
	187.0	188.01	186.84	185.35	187.76
MBNZAM10	287.0	285.56	282.53	285.62	289.69
FEQYUW	183.1	185.90	179.62	181.37	181.45
	342.7	343.06	343.95	338.90	340.34
LILEUC10	182.6	190.35	184.70	185.61	187.98
	347.1	354.47	355.32	353.68	353.08
PHALNC01	353.5	352.05	358.63	357.75	356.82
	178.8	185.52	184.67	184.54	184.71
CYSCLM11	174.9	191.50	179.10	182.07	184.50
	353.5	348.48	357.61	352.42	355.20
TPEPHO02	45.0	34.25	38.65	35.53	34.28

4 Benzoic acid

Here we report the predicted carboxyl ^{13}C and ^{17}O isotropic chemical shifts for both configurations A and B of benzoic acid (Table S10). The values given in Table 5 of the main document are obtained by applying the Boltzmann weighting procedure outlined in Section 7.3 to the values in Table S10.

Table S10: Predicted carboxyl ^{13}C and ^{17}O isotropic chemical shifts for benzoic acid configurations A and B. The data is reported in ppm using fragment, cluster/fragment and GIPAW methods. The absolute shifts were scaled using the linear regression parameters provided in Table 1.

	Config. A			**Config. B**		
	^{13}C: C=O	^{17}O: OH	^{17}O: C=O	^{13}C: C=O	^{17}O: OH	^{17}O: C=O
Fragment:						
OPBE	175.7	173.1	271.6	175.0	173.1	281.7
PBE	175.6	168.5	267.8	174.8	169.2	278.8
TPSS	177.0	170.3	269.6	176.3	170.7	280.3
TPSSh	177.0	170.2	270.2	176.2	170.7	280.6
PBE0	175.7	168.9	269.8	174.9	169.6	279.9
B3LYP	176.0	167.8	268.4	175.2	168.6	278.9
Cluster/Fragment:						
OPBE	175.4	172.6	281.2	175.2	171.9	285.5
PBE	175.2	168.6	278.6	174.9	168.2	283.5
TPSS	176.7	170.1	280.0	176.4	169.6	284.7
TPSSh	176.6	170.1	280.1	176.3	169.7	284.8
PBE0	175.3	169.2	279.3	174.9	168.8	284.0
B3LYP	175.6	168.0	278.8	175.2	167.7	283.5
GIPAW:						
PBE	173.8	167.8	280.8	173.8	169.9	280.8

5 Conversions between chemical shift scales

Conversions between various chemical shift scales can be accomplished based on the IUPAC recommendations:[2]

Abbreviations: Magic-angle spinning (MAS), zero-angle spinning (ZAS)

^{1}H referencing: ^{1}H chemical shifts in the main text are reported on the $\delta_{MAS}^{neat\ TMS}$ scale. On this scale, the chemical shift of solid adamantane under MAS is $\delta_{MAS}^{neat\ TMS}$(adamantane,MAS) = 1.87 ppm. To convert to other chemical shift scales:[2]

$$\delta_{MAS}^{neat\ TMS} = \delta_{MAS}^{1\%\ TMS\ in\ CDCl_3} + 0.124 \text{ ppm}$$

For example, on the 1% TMS in $CDCl_3$ scale, $\delta_{MAS}^{1\%\ TMS\ in\ CDCl_3}$(adamantane,MAS) = 1.75 ppm.

^{13}C referencing: ^{13}C chemical shifts in the main text are reported on the $\delta_{MAS}^{neat\ TMS}$ scale. On this scale, the downfield chemical shift of solid adamantane under MAS is $\delta_{MAS}^{neat\ TMS}$(adamantane,MAS) = 38.48 ppm. To convert to other chemical shift scales:[2, 3]

$$\begin{aligned}
\delta_{MAS}^{neat\ TMS} &= \delta_{ZAS}^{neat\ TMS} \\
&= \delta_{MAS}^{1\%\ TMS\ in\ CDCl_3} + 0.71 \text{ ppm} \\
&= \delta_{MAS}^{0.5\%\ DSS\ in\ D_2O} - 2.01 \text{ ppm} \\
&= \delta_{ZAS}^{0.5\%\ DSS\ in\ D_2O} - 2.67 \text{ ppm}
\end{aligned}$$

^{15}N referencing: ^{15}N chemical shifts in the main text are reported on the $\delta_{MAS}^{solid\ NH_4Cl}$ scale. To convert to other chemical shift scales:[2, 4]

$$\delta_{MAS}^{solid\ NH_4Cl} = \delta_{MAS}^{neat\ MeNO_2} + 341.17 \text{ ppm}$$

$$\delta_{MAS}^{liq-NH_3} = \delta_{ZAS}^{neat\ MeNO_2} + 380.44 \text{ ppm}$$

6 Sample Quantum Espresso geometry optimization job

This section provides a sample Quantum Espresso geometry optimization job which mimics the procedures used in the development of the test set. Crystal structures optimized with fixed lattice parameters and this protocol should be particularly well-suited for use with the reported scaling parameters.

```
 &CONTROL
                 calculation = 'relax',
                restart_mode = 'from_scratch' ,
                  pseudo_dir = 'upf_files' ,
                      prefix = 'glycine' ,
                     disk_io = 'default' ,
                   verbosity = 'high' ,
               etot_conv_thr = 1.0D-4 ,
 /
 &SYSTEM
                       ibrav = 0,
                         nat = 40,
                        ntyp = 4,
                     ecutwfc = 80 ,
                     ecutrho = 320 ,
            vdw_corr = 'DFT-D',
 /
 &ELECTRONS
                    conv_thr = 1.D-8 ,
                mixing_beta = 0.5D0 ,
 /
 &IONS
 ion_dynamics = "bfgs"
 /

ATOMIC_SPECIES
 H 1.00000 H.pbe-rrkjus.UPF
 C 12.0000 C.pbe-rrkjus.UPF
 N 14.0000 N.pbe-rrkjus.UPF
 O 16.0000 O.pbe-rrkjus.UPF

ATOMIC_POSITIONS (crystal)
H 0.1043 0.8841 0.3413
O 0.1478613 0.858457 0.8928313
H 0.9263 0.7771 0.2423
C 0.9355415 0.854986 0.2130813
N 0.6988114 0.910176 0.2589512
H 0.7154 0.8991 0.4453
H 0.5173 0.8841 0.1433
H 0.7064 0.9881 0.2313
O 0.6941413 0.905676 0.7643110
```

```
C 0.9246914 0.875196 0.9335313
O 0.3521413 0.358457 0.6071713
O 0.8058613 0.405676 0.7356910
C 0.5753114 0.375196 0.5664713
H 0.5743 0.2771 0.2583
H 0.3963 0.3841 0.1593
C 0.5644615 0.354986 0.2869213
N 0.8011914 0.410176 0.2410512
H 0.7854 0.3991 0.0553
H 0.9833 0.3841 0.3573
H 0.7944 0.4881 0.2693
O 0.3058613 0.094336 0.2356910
C 0.0753114 0.124816 0.0664713
H 0.0743 0.2231 0.7583
C 0.0644615 0.145026 0.7869213
N 0.3011914 0.089836 0.7410512
H 0.2854 0.1011 0.5553
H 0.4833 0.1161 0.8573
H 0.2944 0.0121 0.7693
O 0.8521413 0.141557 0.1071713
H 0.8963 0.1161 0.6593
O 0.6478613 0.641557 0.3928313
O 0.1941413 0.594336 0.2643110
C 0.4246914 0.624816 0.4335313
H 0.4263 0.7231 0.7423
H 0.6043 0.6161 0.8413
C 0.4355415 0.645026 0.7130813
N 0.1988114 0.589836 0.7589512
H 0.2154 0.6011 0.9453
H 0.0173 0.6161 0.6433
H 0.2064 0.5121 0.7313


K_POINTS automatic
3 3 3 1 1 1

CELL_PARAMETERS angstrom
5.10470000    0.00000000    0.00000000
0.00000000   11.97200000    0.00000000
-2.02350669    0.00000000    5.07453271
```

7 Details of NMR experiments performed here

Figure S9 illustrates the the experimental ^{15}N solid state NMR spectra for structures 8, and 18–27 in Figure S2. The specific chemical shifts are listed in Table S6.

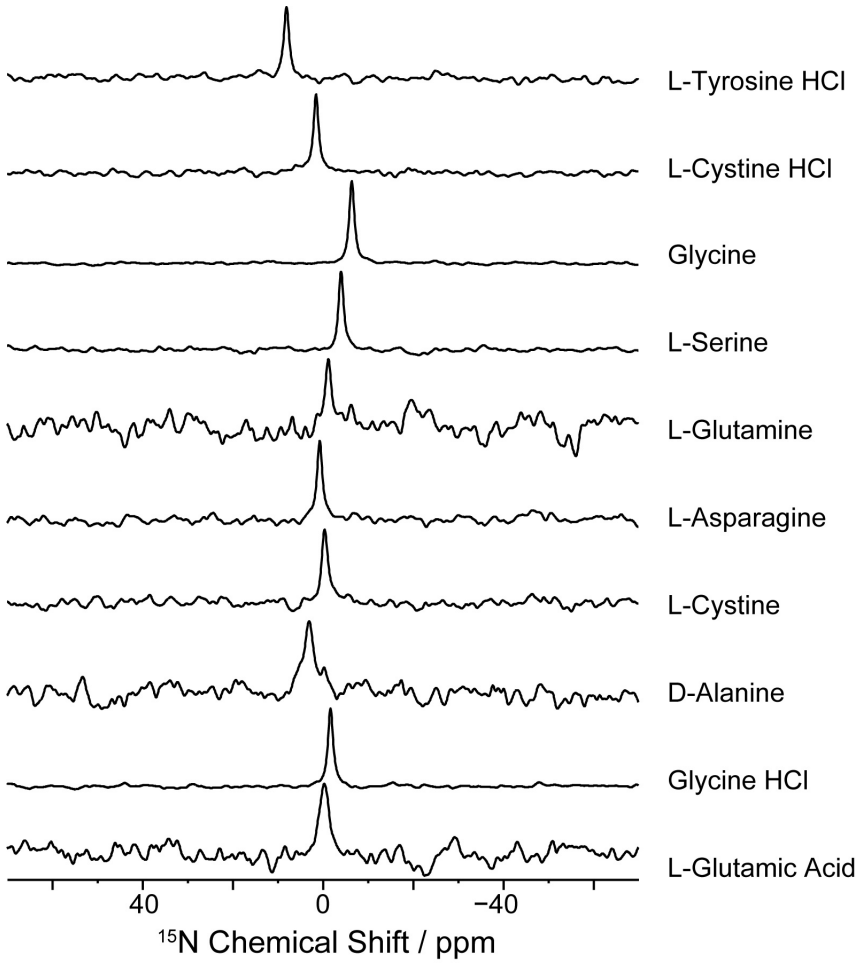


Figure S9: Experimental ^{15}N solid state NMR spectra for structures 8, and 18–27 in the nitrogen test set.

References

[1] Hartman, J.D., Monaco, S., Schatschneider, B., and Beran, G.J.O. "Fragment-based ^{13}C nuclear magnetic resonance chemical shift predictions in molecular crystals: An alternative to planewave methods." *J. Chem. Phys.*, **143**, 102809 (2015).

[2] Harris, R.K., Becker, E.D., Cabral de Menezes, S.M., Granger, P., Hoffman, R.E., and Zilm, K.W. "Further conventions for NMR shielding and chemical shifts (IUPAC Recommendations 2008)." *Pure Appl. Chem.*, **80**, 59–84 (2008). doi:10.1351/pac200880010059.

[3] Morcombe, C.R. and Zilm, K.W. "Chemical shift referencing in MAS solid state NMR." *J. Magn. Res.*, **162**, 479–486 (2003). doi:10.1016/S1090-7807(03)00082-X.

[4] Hayashi, S. and Hayamizu, K. "Chemical shift standards in high-resolution solid-state NMR (2) ^{15}N nuclei." *Bull. Chem. Soc. Jpn.*, **64**, 688–690 (1991).