Electronic Supplementary Material (ESI) for Faraday Discussions.

Supplementary information: The essential synergy of MD simulation and NMR in understanding amorphous drug forms

Jamie L. Guest, Esther A. E. Bourne, Martin A. Screen, Mark R. Wilson, Tran N. Pham and Paul Hodgkinson

1. Molecular dynamics simulation

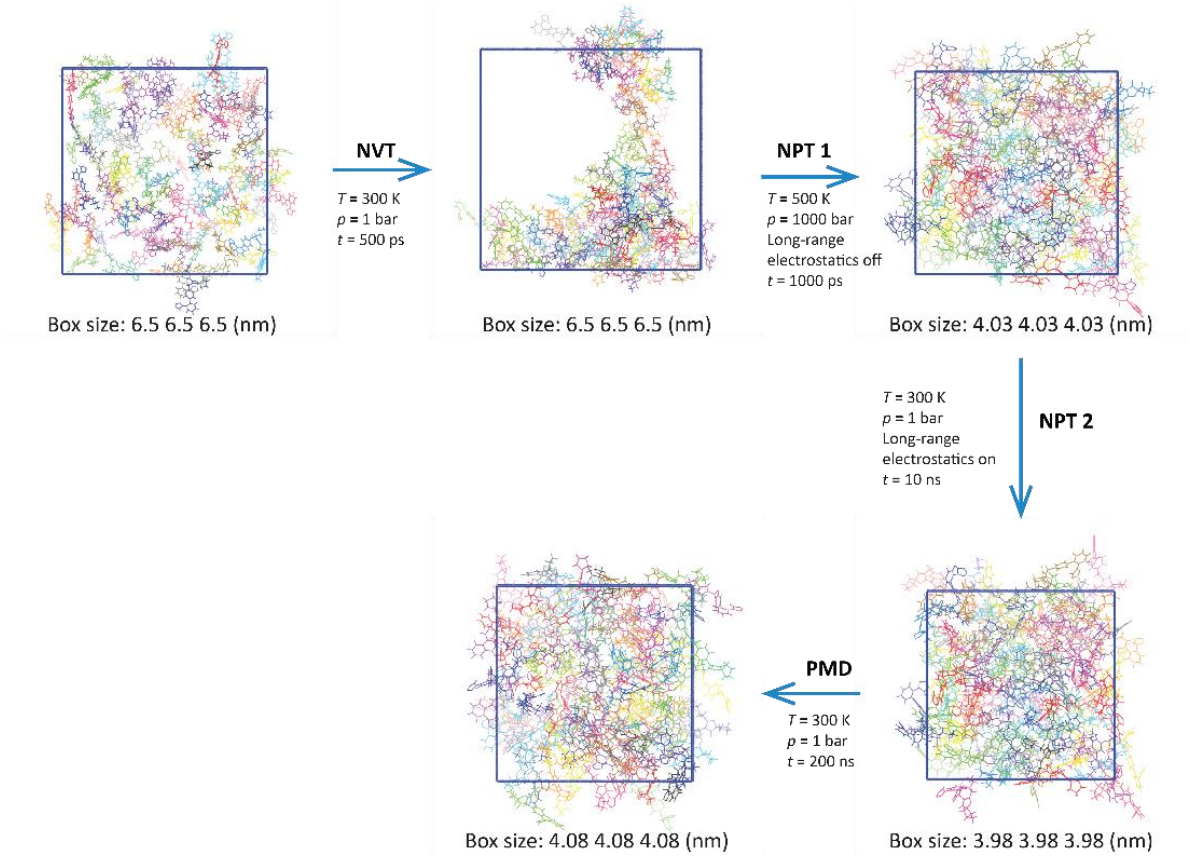


Figure S1 Schematic illustration of the MD protocol used to generate the amorphous glasses of IRB. See the main text for the details of each simulation step. PMD refers to the final 200 ns production run. Note that the exact box dimensions following compression (NPT1 step) vary; the figures shown here are for glass 1 of 1H tautomer.

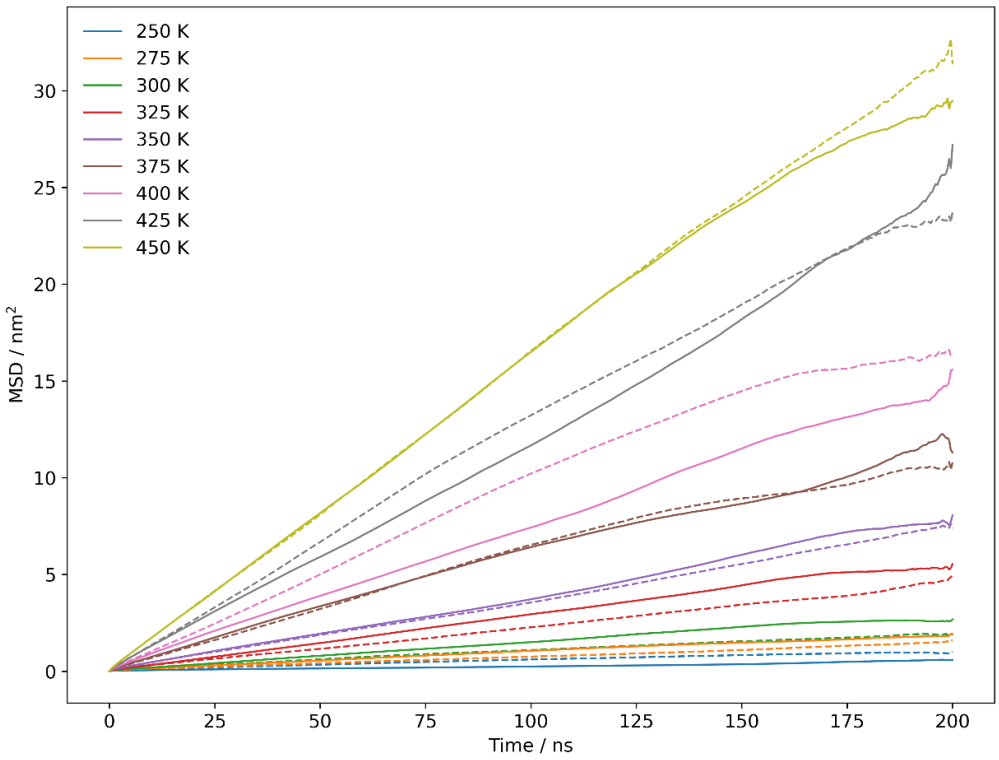


Figure S2 Mean squared displacement (MSD) averaged over all C11 atoms in the simulation as a function of simulation time for IRB-1H at different temperatures, calculated using the GROMACS `msd` command. Repeats for each temperature are shown in the same colour, with a solid line for the heating run and a dashed line for the cooling run. The diffusion coefficients in Figure 1 are derived from linear regression of this data between 10–175 ms.

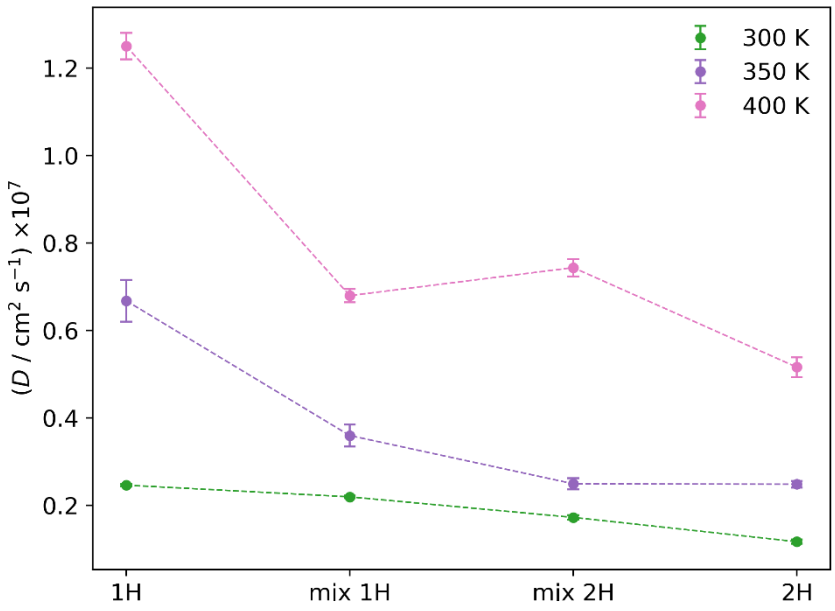


Figure S3 Mean diffusion coefficients calculated from the MSD averaged over all C11 atoms for IRB molecules in pure IRB-1H and IRB-2H, and a 50:50 mixture of tautomers (e.g. “mix 1H” refers to the 1H molecules in the mixture). Data is shown for the heating simulations only and at selected temperatures for clarity, and the dashed lines are guides to the eye.

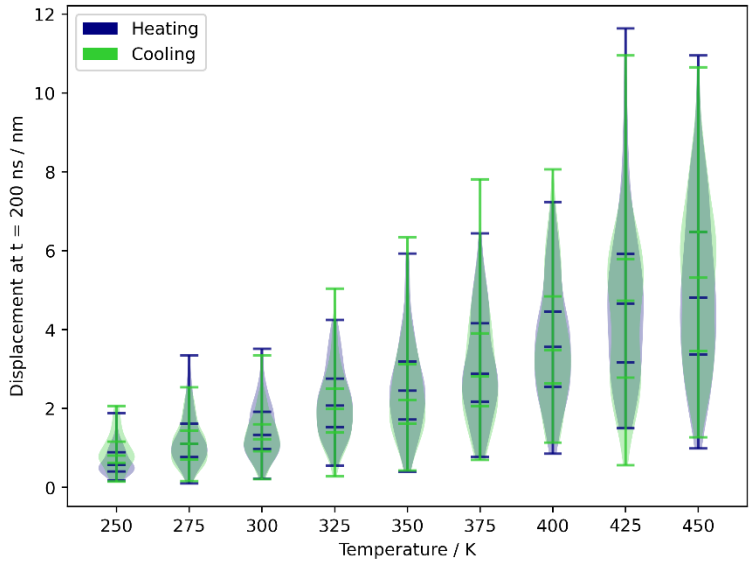


Figure S4 Distributions of the final displacement of each IRB molecule (using C11 as the reference atom) in simulations of pure IRB-1H at different temperatures, and for both heating and cooling runs. Horizontal markers show the range of the data and the median, upper, and lower quartiles.

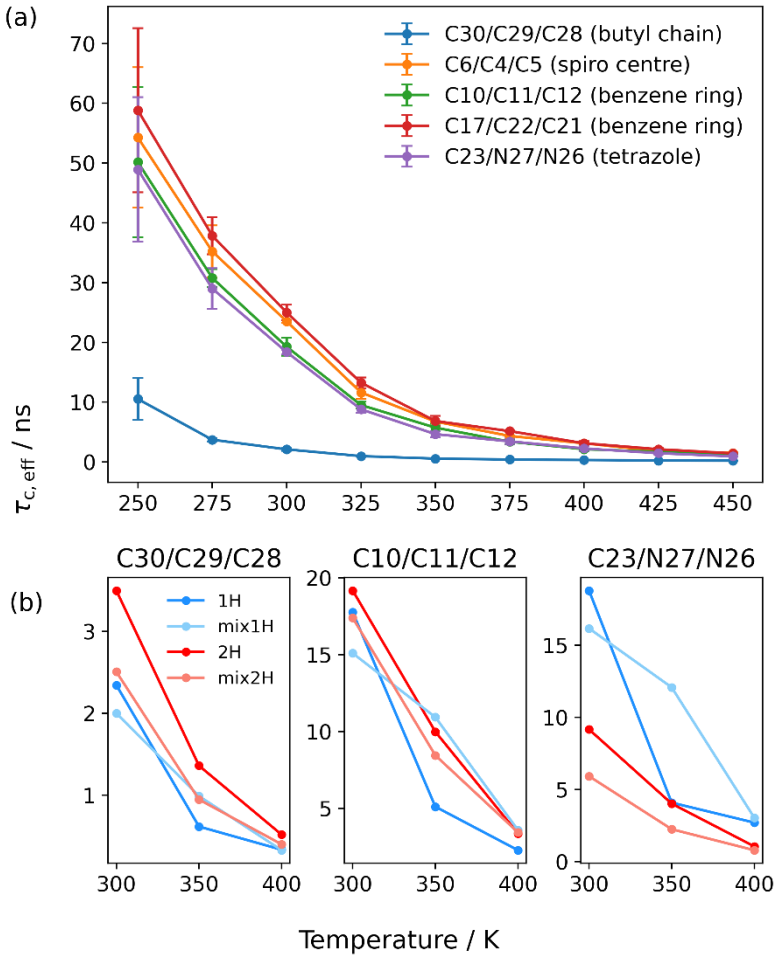


Figure S5 (a) Effective correlation times for characteristic vectors in different parts of the IRB molecule as a function of temperature for pure IRB-1H, averaged over the heating and cooling runs. The vectors are cross-products between interatomic vectors between a central atom and its two neighbours, with the atom

numbering of Scheme 1. (b) Comparison between tautomer simulations for selected regions of the IRB molecule (butyl chain, a benzene ring and tetrazole ring) for simulations of pure IRB-1H and IRB-2H, and the 50:50 mixture (1H and 2H components of this mixture are labelled mix1H and mix2H respectively).

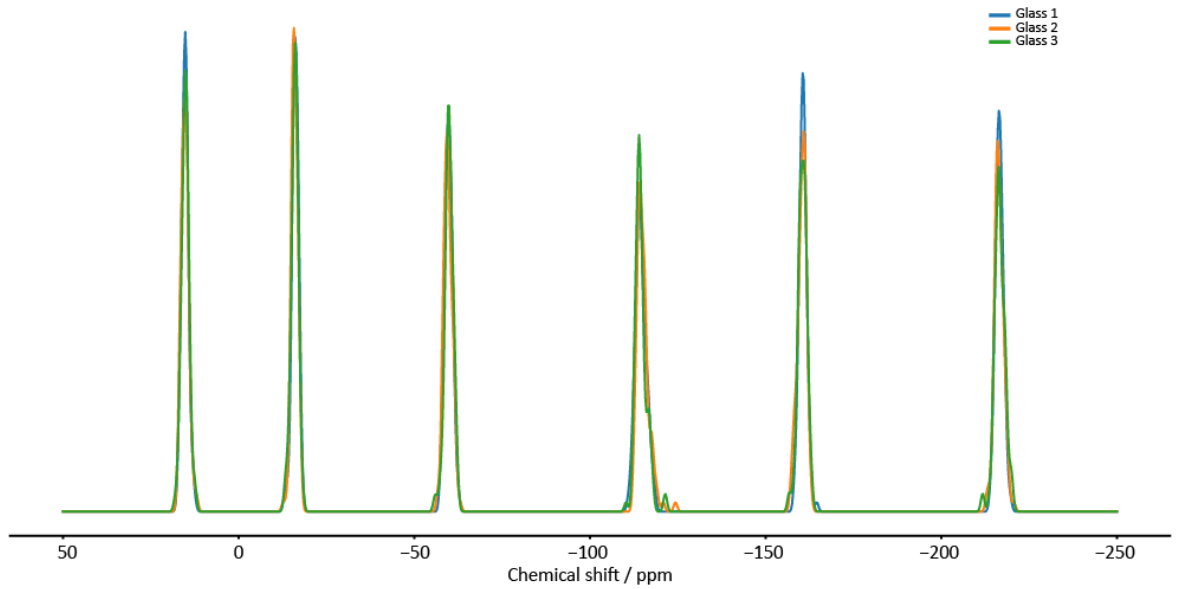


Figure S6 ShiftML2-predicted ^{15}N spectra calculated from three models of amorphous 1H irbesartan generated by the MD protocol above (shifts averaged over the production run). The overall predictions are seen to be highly reproducible.

Table S1 Average uncertainties (in ppm) for ShiftML2 predictions of C isotropic shieldings in the irbesartan tautomers.

	C10	**C11**	**C12**	**C13**	**C14**	**C15**	**C16**	**C17**	**C18**	**C19**	**C2**	**C20**	**C21**
1H	3.76	4.10	2.44	2.57	4.02	2.56	2.46	3.70	4.81	2.46	7.71	1.93	1.90
2H	3.74	4.08	2.45	2.40	3.56	2.41	2.44	3.57	4.78	2.49	7.65	1.92	1.89
	C22	**C23**	**C28**	**C29**	**C30**	**C31**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	
1H	2.33	7.01	3.43	2.66	2.46	2.07	8.26	7.07	3.26	2.52	2.51	3.25	
2H	2.26	6.82	3.45	2.67	2.47	2.06	8.29	7.15	3.26	2.54	2.54	3.25	

Table S2 Average uncertainties (in ppm) for ShiftML2 predictions of N isotropic shieldings in the irbesartan tautomers

	N1	**N24**	**N25**	**N26**	**N27**	**N3**
1H	9.94	12.40	9.64	8.82	10.63	15.05
2H	9.81	12.32	15.13	10.91	11.43	14.98

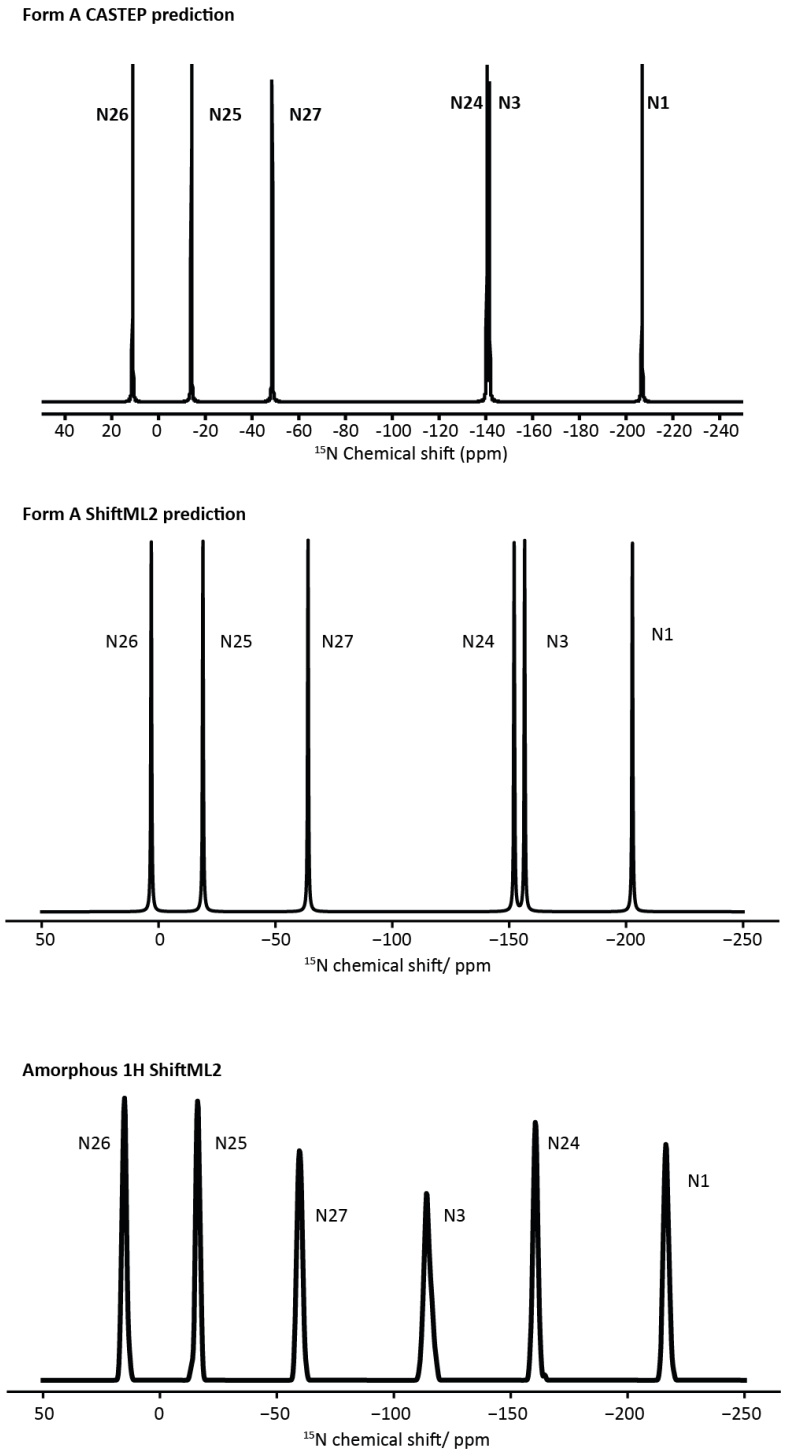


Figure S7 CASTEP (upper) and ShiftML2 (middle) predicted ^{15}N NMR spectra of IRB form A (which contains the 1H tautomer only) compared to the ShiftML2 shift prediction of the amorphous 1H tautomer (lower), constructed via MD simulation.

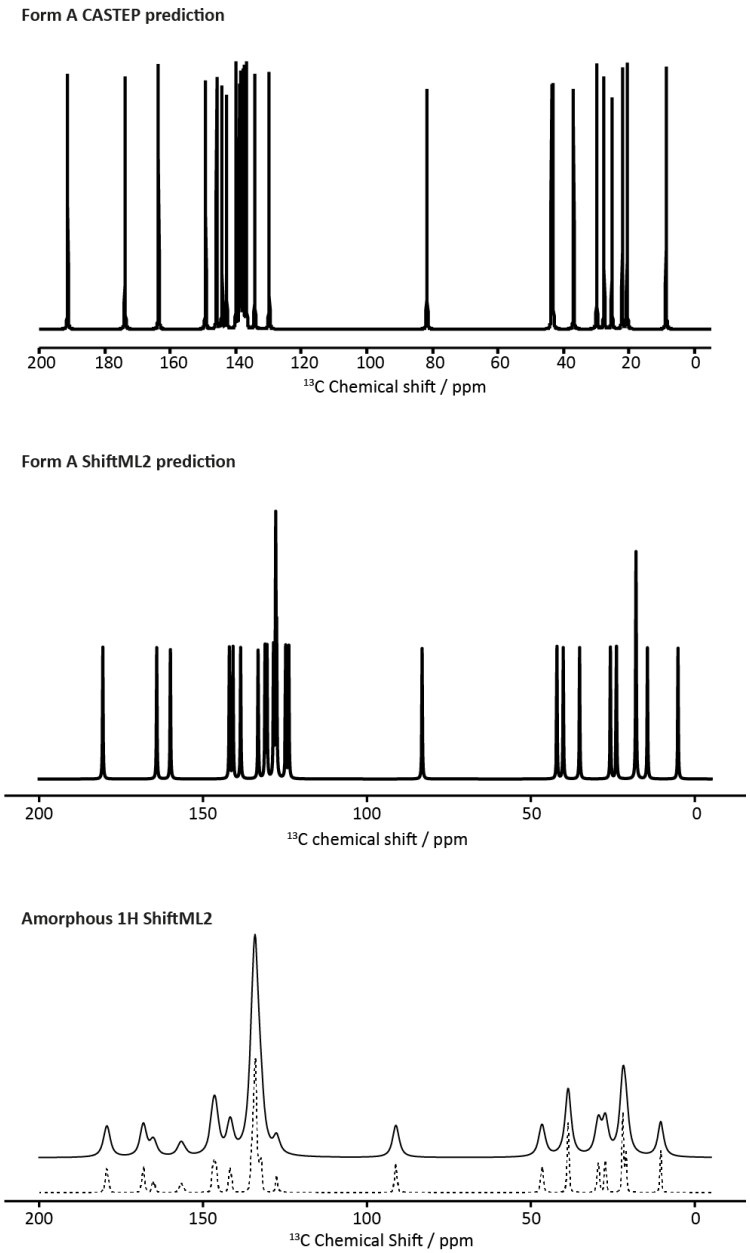


Figure S8 CASTEP (upper) and ShiftML2 (middle) ^{13}C chemical shift prediction of IRB form A compared to the ShiftML2 shift prediction of the amorphous 1H tautomer (lower), constructed via MD simulation. The amorphous spectrum was convoluted with Lorentzian peaks with two different line widths: FWHM of 2 ppm (solid line) and 0.2 ppm (dotted).

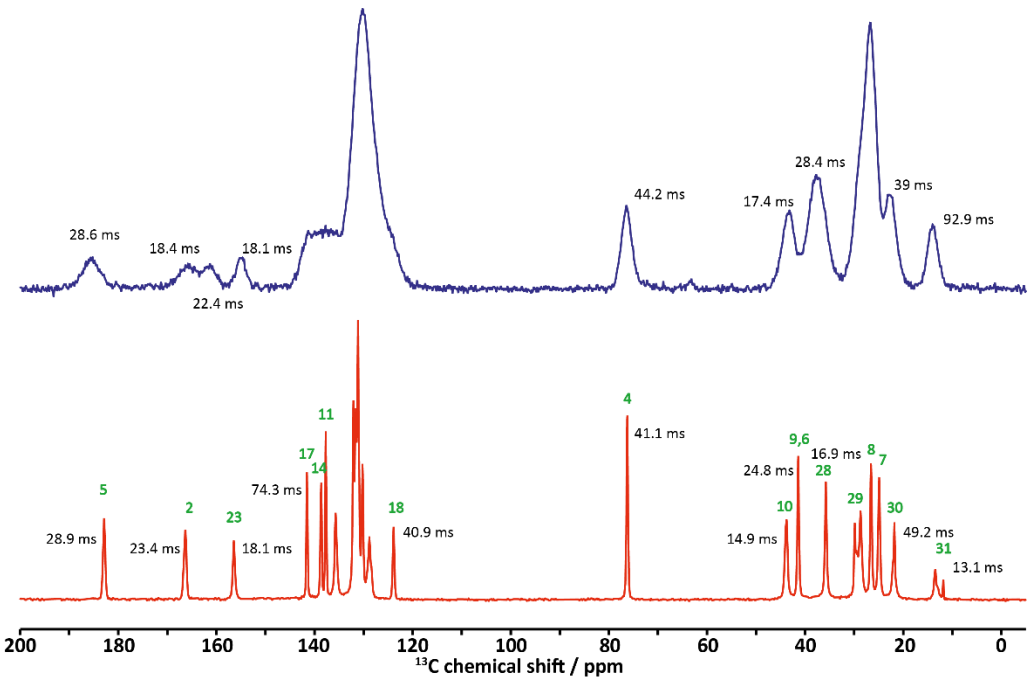
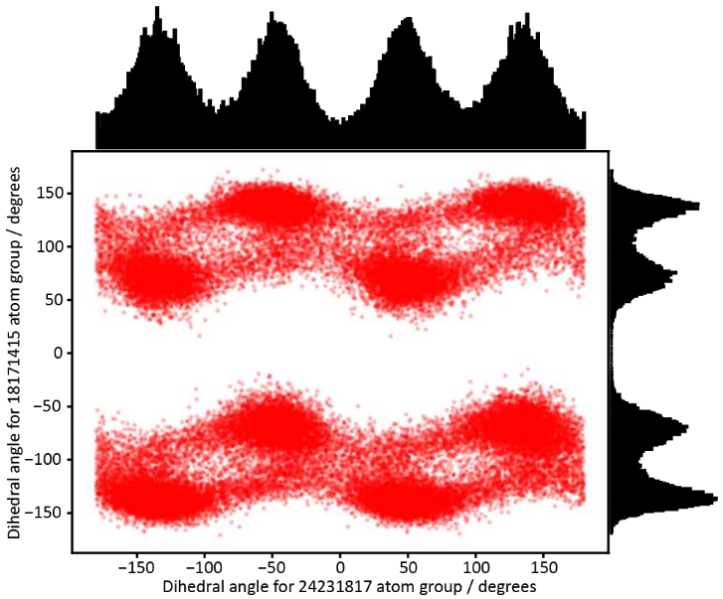


Figure S9 ^{13}C CP/MAS spectra of (top) amorphous and (bottom) form A of irbesartan, annotated with T_2' time constants for the resolved peaks. The T_2' values were obtained by performing a spin echo on carbon. The pulse sequence was first calibrated, and the SPINAL64 decoupling was optimised, on a sample of glycine to achieve a T_2' of 25 ms on C_α. Site labels (green) follow Scheme 1 of the main text.



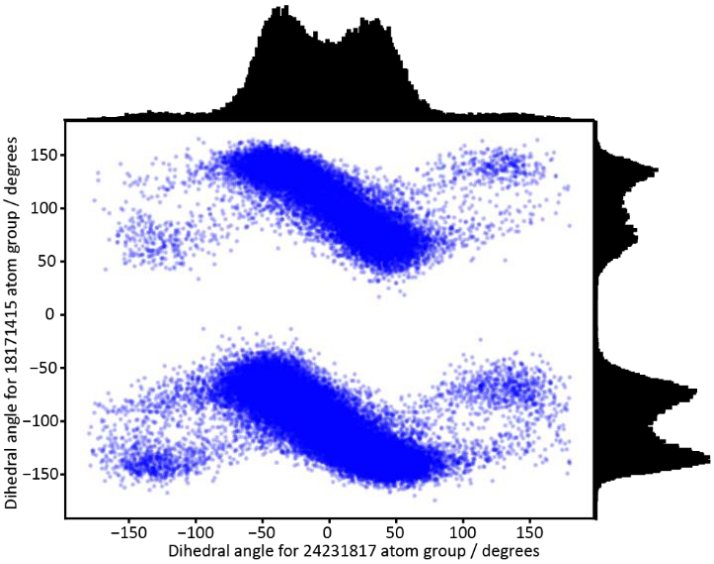


Figure S10 Plot showing the correlation between two torsion angles corresponding to relative ring orientations (C18-C17-C14-C15 vs N24-C23-C18-C17) in 2H (upper) and 1H (lower) tautomers of irbesartan, for every snapshot and molecule of the MD simulations. The horizontal and vertical axis projections are histograms of the frequency distributions of the individual torsion angles.

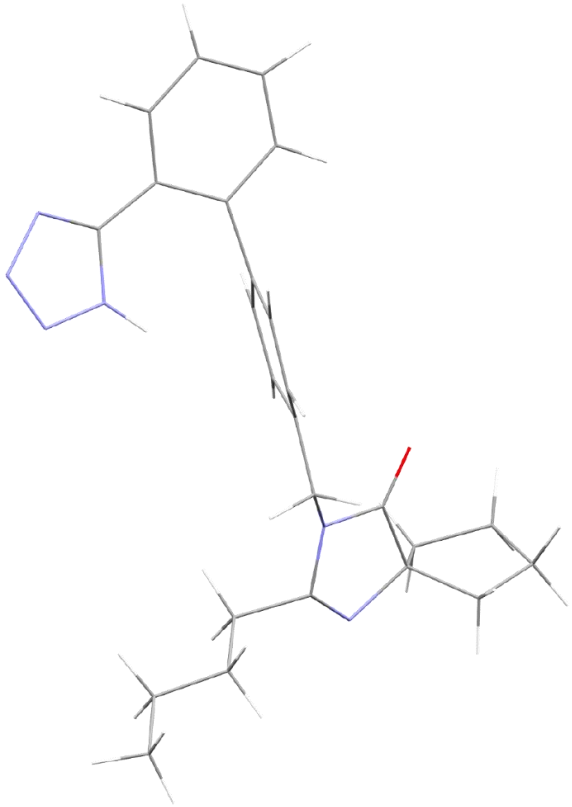


Figure S11 Example conformation of IRB 1H in which the tetrazole ring hydrogen (top left of figure) is pointing towards the central phenyl ring. The N24-C23-C18-C17 and C18-C17-C14-C15 torsion angles are –0.45° and –104.3° respectively in this example.

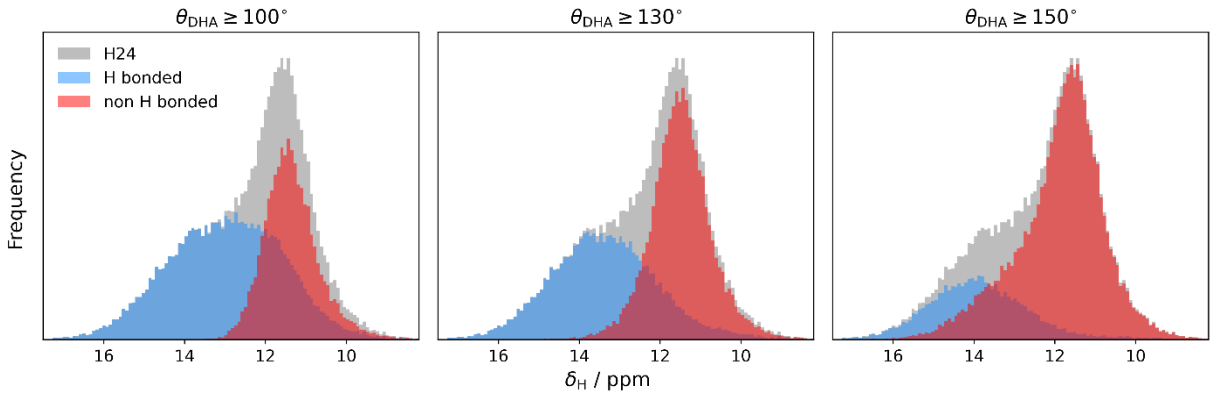


Figure S12 ShiftML2-predicted unique chemical shifts of all tetrazole H atoms from 501 frames of MD simulations of IRB-1H at 300 K. Three example data sets are shown where each H is categorised as hydrogen bonded or non-hydrogen bonded, using different angular cutoffs from 100° to 150°, at a constant donor-acceptor distance of 3.5 Å.

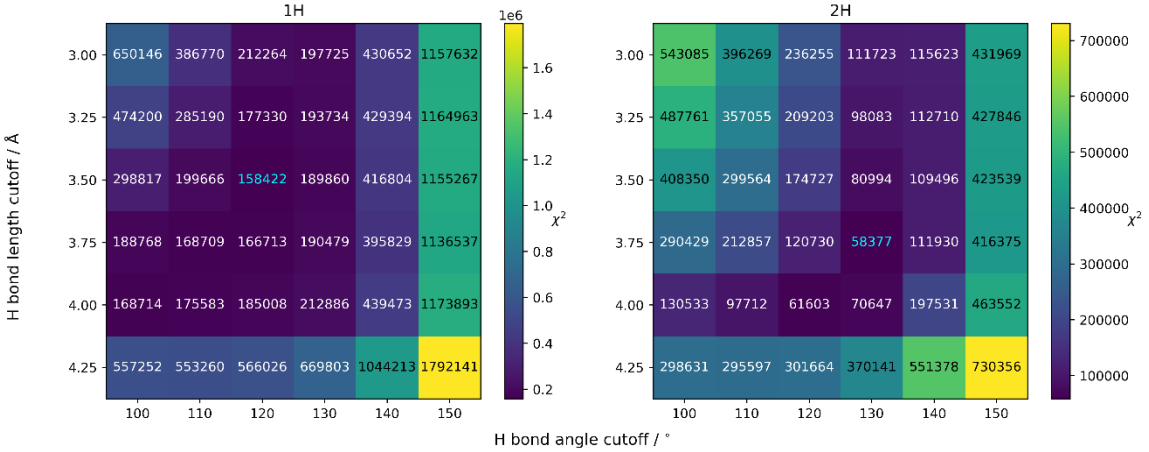


Figure S13 χ^2 values for different hydrogen bond angle and length cutoffs tested for IRB-1H and IRB-2H, with the smallest value for each set highlighted in blue. The chemical shifts predicted for every tetrazole H site and in each snapshot were separated into hydrogen bonded and non hydrogen bonded categories and these two distributions were separately fitted to a single gaussian curve. These two curves for each tautomer were then summed to give an overall fit for the chemical shift data for H24/H25, with the deviation between this fit and the data used to calculate χ^2.

The χ^2 values as a function of r_{DA} and θ_{DHA} shown in Fig. S12 indicate that the angular cutoff has a greater influence than the distance on the number of hydrogen bonds identified; there is significantly more change in χ^2 as θ_{DHA} is varied for a constant r_{DA} than vice versa. It is clear that for both tautomers $\theta_{HDA} \geq 150°$ is too restrictive and is likely to underestimate the number of hydrogen bonds, and $r_{DA} \leq 4.25$ Å is too lenient and will predict too many. The choices of r_{DA} and θ_{HDA} that fit each data set best do not differ significantly, and the final values used for the remaining hydrogen bonding analysis were chosen as $\theta_{HDA} \geq 125°$ and $r_{DA} \leq 3.5$ Å, with the angular cutoff midway between the optimal value for each tautomer, and the shorter distance cutoff of the two.

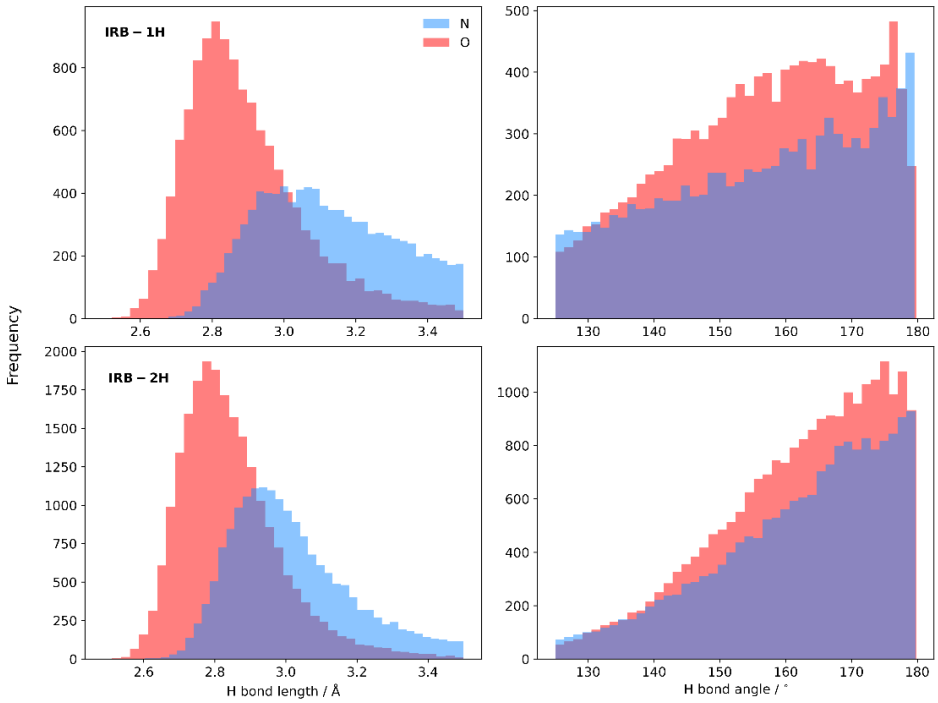


Fig. S14 Distributions of hydrogen bond lengths (donor-acceptor distance) and angles for single hydrogen bonds only in simulations of IRB-1H and IRB-2H at 300 K, with bonds to N atoms shown in blue, and O in red. An angular cutoff of 125° and a distance cutoff of 3.5 Å were used, and a volume element correction factor has been applied to the bond angle distributions to remove the dependence on $\sin\theta_{\mathrm{HDA}}$.

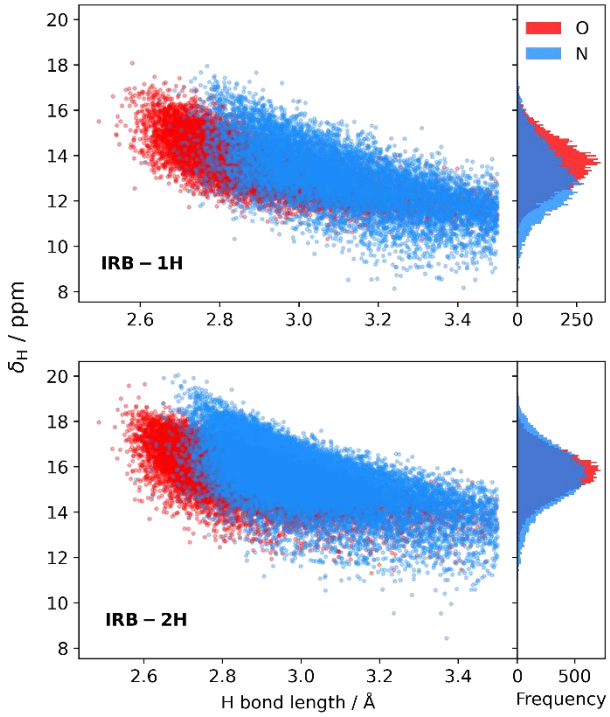


Figure S15 Tetrazole hydrogen bond lengths (donor—acceptor distances) for single hydrogen bonds only in each snapshot of simulations of IRB-1H and IRB-2H at 300 K, plotted against the chemical shift of the tetrazole H. Hydrogen bonds where the acceptor is O are shown in red, and N in blue. Note how the overall correlations and the mean shifts differ between tautomers.

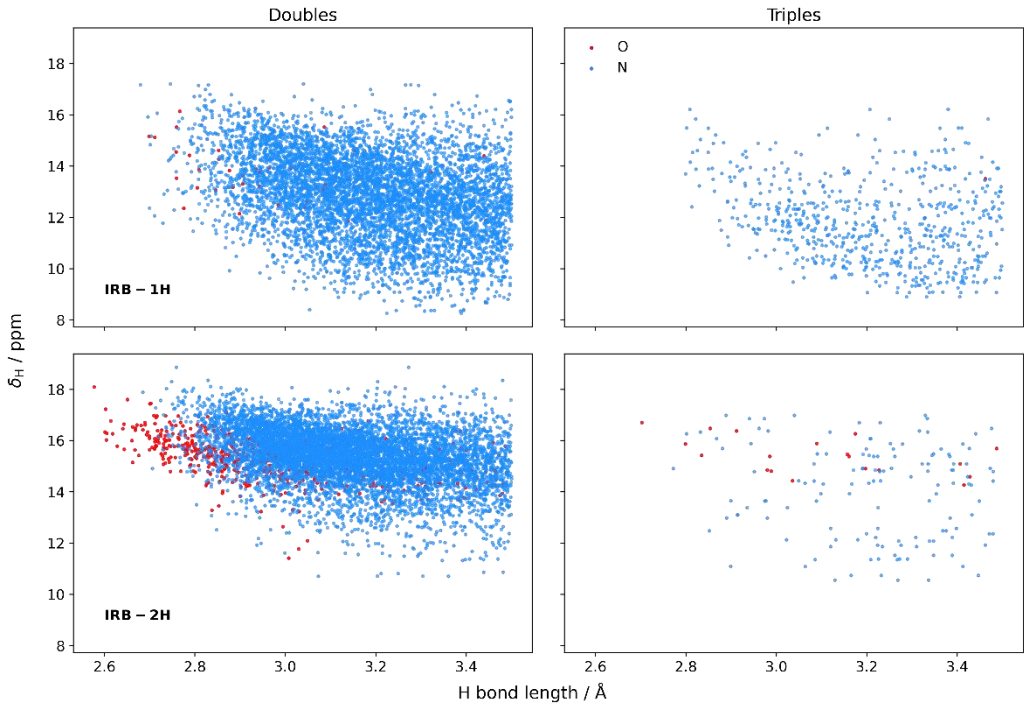


Figure S16 Tetrazole hydrogen bond lengths (donor-acceptor distances) for "complex" hydrogen bonds in each snapshot of simulations of IRB-1H and IRB-2H at 300 K, plotted against the chemical shift of the tetrazole H. Hydrogen bonds with 2 acceptor atoms are shown on the left, and those with 3 on the right. Each bond length between H24/H25 and its acceptor atoms is plotted against the shared chemical shift value.

Table S3 Total number of hydrogen bonds identified from the 501 snapshots of simulations of IRB-1H and IRB-2H at 300 K, using a maximum donor-acceptor distance of 3.5 Å and a minimum donor-hydrogen-acceptor angle of 125°.

	IRB-1H			**IRB-2H**		
	Singles	Doubles	Triples	Singles	Doubles	Triples
Counts	21780	3008	211	39990	3772	51
NH—O[a]	12313	44	1	22253	592	19
NH—N[a]	9467	5972	632	17737	6952	134

[a] The NH—O and NH—N rows list each occurrence of the respective acceptor atoms O and N participating in each bond, for example a double hydrogen bond shared between two different N acceptors is recorded as 1 double bond, with 2 counts in the NH—N row

Of the 21780 single hydrogen bonds identified for the IRB-1H simulation, 85 are intramolecular, with O as the acceptor atom in all cases. In the IRB-2H simulation, only are 6 intramolecular hydrogen bonds were observed, 5 with N3 as the acceptor and 1 with O.

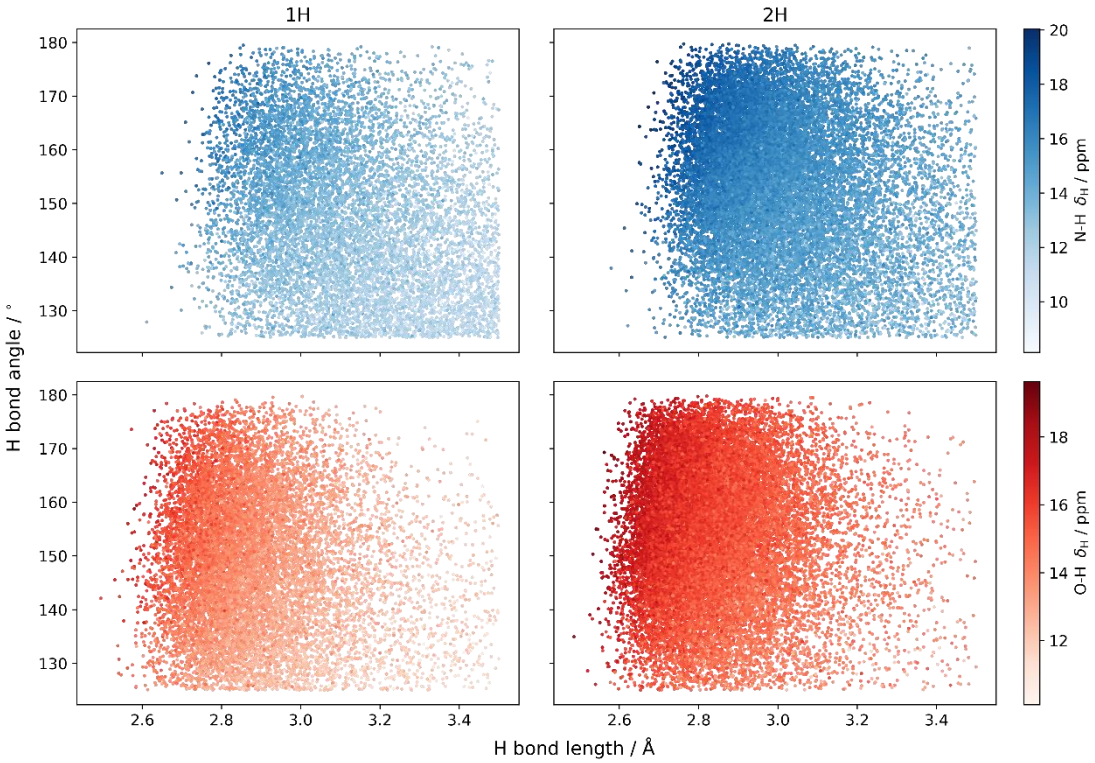


Figure S17 Hydrogen bond lengths (donor-acceptor distances) and angles for single hydrogen bonds of the tetrazole H to N and O in simulations of IRB-1H and IRB-2H at 300 K, against its chemical shift. N—H bonds are shown in blue and O—H bonds in red. The maximum donor-acceptor distance and minimum donor-hydrogen-acceptor angle were set to 3.5 Å and 125° respectively.

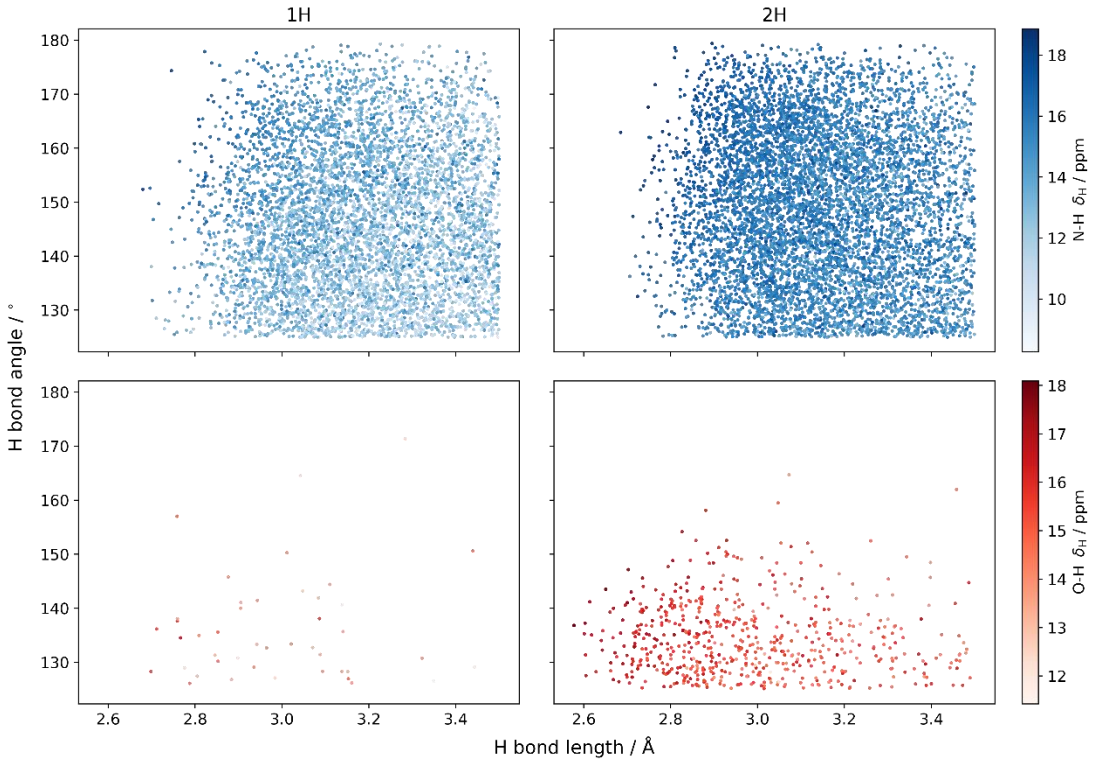


Figure S18 Hydrogen bond lengths and angles for double hydrogen bonds (where there are 2 acceptor atoms) of the tetrazole H to N and O in simulations of IRB-1H and IRB-2H at 300 K, against its chemical shift. The maximum donor-acceptor distance and minimum donor-hydrogen-acceptor angle were set to 3.5 Å and 125° respectively.

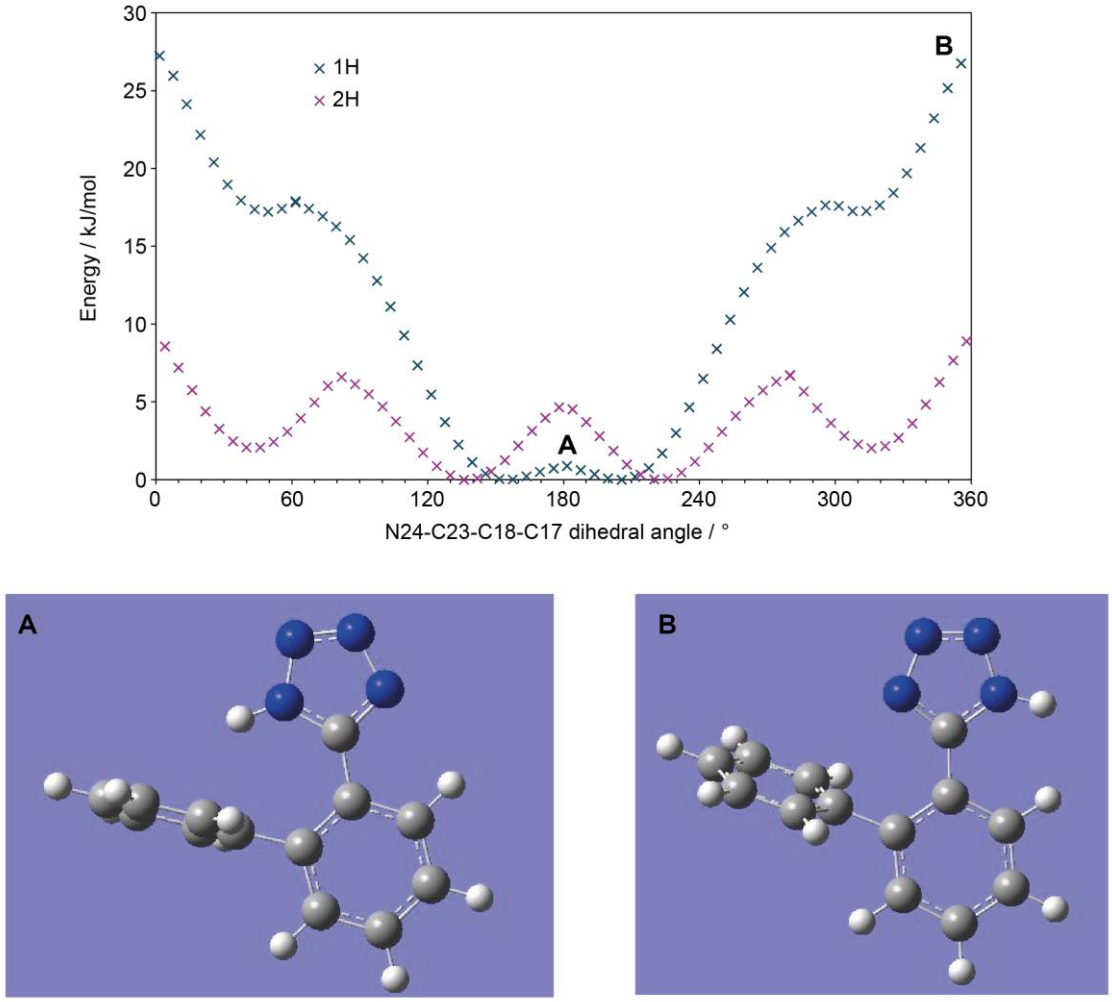


Figure S19 Dihedral angle energies for molecular fragments of 1H and 2H tautomers. Energies are calculated at the B3LYP//6-311+g(d,p) level using Grimme's D3 dispersion corrections using the Gaussian 16 program. Structure A shows the lowest conformation for 1H, where there is a weak interaction between a hydrogen and the phenyl ring. Structure B shows the highest energy structure for 1H, corresponding to a steric clash between the same hydrogen and a hydrogen on the phenyl ring.