Supporting Information for

# An Efficient Strategy for Determining the Atomic-Resolution Structure of Micro- and Nanocrystalline Solids within Polymeric Microbeads: Domain-Edited NMR Crystallography

Jiri Brus, \* Jiri Czernek, Martin Hruby, Pavel Svec, Libor Kobera, Sabina Abbrent, and Martina Urbanova

Institute of Macromolecular Chemistry, Czech Academy of Sciences, Heyrovsky sq. 2, 162 06 Prague 6, Czech Republic.

AUTHOR EMAIL ADDRESSES: brus@imc.cas.cz, czernek@imc.cas.cz, mhruby@centrum.cz, prion2@seznam.cz, kobera@imc.cas.cz, abbrent@imc.cas.cz, urbanova@imc.cas.cz

Table of Content

**Supporting Information SI1 – Synthesis** 

Supporting Information SI2 – Solid-State NMR Spectroscopy

Supporting Information SI3 – Computational procedures

Supporting Information SI4 – Long-range <sup>1</sup>H-<sup>1</sup>H correlation spectroscopy

Supporting Information SI5 – Isotropic values of <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N NMR chemical shifts of Decitabine Polymorphic Form X (DAC-X)

Supporting Information SI6 - Similarity measures for DAC-I (reference) and DAC-X

## **Supporting Information SI1 – Synthesis**

*Materials:* Decitabine (5-aza-2'-deoxycytidine, DAC) were purchased from Sigma-Aldrich Czech Republic.

Synthesis of poly(sebacic acid-co-1,4-cyclohexanedicarboxylic acid) (70:30 mol/mol) (PSAco-PCH): Sebacic acid (8.40 g, 41.5 mmol), 1,4-cyclohexanedicarboxylic acid (3.06 g, 17.8 mmol) and acetic anhydride (115 mL, 124 g, 1210 mmol) were refluxed for 30 min, and the reaction mixture was evaporated in vacuo. Subsequently, the residues of acetic acid and acetic anhydride were azeotropically removed by evaporation with toluene (2x), and the solid residue was dissolved in chloroform, followed by precipitation with diethyl ether – petroleum ether mixture (1:1 v/v). The precipitated pre-polymer was then filtered off, air-dried, and heated on magnetic stirrer at 180 °C for 90 min in vacuo (10 Pa), and cooled. Finally, the solidified melt was dissolved in dichloromethane and precipitated with petroleum ether, yielding 7.79 g (75%) of purified PSA-co-PCH. Weight-average molecular weight (Mw): 12.1 kDa, according to gel permeation chromatography.

*Microbeads preparation:* Subsequently, powdered DAC was ground in acetonitrile with an IKA T25 Ultra Turrax® dispersing instrument (Fisher Scientific Ltd, Pardubice, Czech Republic) to form a suspension, filtered off, dried on air and milled in dry state with a Pulverisette 23® Mini Mill (Fritsch, ILABO Ltd, Kyjov, Czech Republic). Fraction > 500 mesh was then sieved out for use. Subsequently PSA-co-PCH (420 mg) was dissolved in anhydrous acetonitrile (1.78 mL) at 60 °C and milled DAC was added. This suspension (58 °C hot) was emulgated into 50 mL of polyisobutylene-thickened mineral oil (58 °C hot) with vigorous stirring with anchor stirrer (2000 rpm) and the emulsion was stirred at ca 60 °C until acetonitrile is completely evaporated (ca 30 min). The suspension was cooled to room temperature while stirring and filtered. The collected beads with average diameter of 125  $\mu$ m were washed several times with hexane from mineral oil and polyisobutylene, dried on air, quickly washed by water to remove surface-bound drug crystals and immediately dried.

## Supporting Information SI2 – Solid-State NMR Spectroscopy

All solid-state NMR spectra were measured at 11.7 T using a Bruker Avance 500 WB/US NMR spectrometer (2013) in a double-resonance 4-mm probe-head at spinning frequencies 10 -11 kHz. For explicit determination of isotropic chemical shifts, the following techniques were used: i) <sup>1</sup>H NMR with DUMBO homodecoupling,<sup>1</sup> ii) <sup>13</sup>C and <sup>15</sup>N CP/MAS and <sup>13</sup>C CPPI/MAS NMR,<sup>2,3</sup> iii) 2D <sup>1</sup>H-<sup>13</sup>C FSLG HETCOR NMR,<sup>4</sup> iv) 2D NOESY-type <sup>1</sup>H-<sup>1</sup>H spin-diffusion NMR with DUMBO homodecoupling,<sup>5</sup> and v) 2D DQ/SQ <sup>1</sup>H-<sup>1</sup>H DUMBO MAS NMR<sup>6</sup> with SPC5 DQ recoupling<sup>7</sup>. To suppress unwanted coherences, the  $T_1(^{1}\text{H})$  filter consisting of 180° (<sup>1</sup>H) pulse followed by a short delay was used. Frictional heating<sup>8,9</sup> of the spinning samples was compensated for by active cooling. For all the experimental details, see sections below.

# 1D MAS NMR experiments:

The <sup>13</sup>C CP/MAS NMR spectra employing cross-polarization were acquired using the standard pulse scheme at spinning frequency of 11 kHz. The recycle delay was 30 s and the cross-polarization contact time was ranging from 0.1 to 3 ms. The strength of spin-locking fields  $B_1(^{13}C)$  expressed in frequency units  $\omega_l/2\pi = \gamma B_1$  was 64 kHz. The <sup>13</sup>C CPPI/MAS NMR spectra employing cross-polarization polarization-inversion (CPPI) preparation period to distinguish CH<sub>n</sub> groups were measured using the standard pulse scheme at spinning frequency of 11 kHz. The cross-polarization contact time was 1.75 ms whereas duration of the polarization inversion period was 60 µs. The spectra were referenced to  $\alpha$ -glycine (176.03 ppm). The number of scans was usually 1600 which corresponds to the total experimental time of ca. 13 hours.

The <sup>1</sup>H MAS NMR spectra with DUMBO homodecoupling were measured at 10 kHz (MAS frequency) and number of scans 64-128. The 90° (<sup>1</sup>H) pulse-length was 2.2  $\mu$ s, power level for DUMBO shape pulse was 71 W, DUMBO pulse length 32  $\mu$ s and the number of loops for digital averaging was 6. All parameters were optimized on glycine to reach maximum spectral resolution ( $\Delta v$ (NH<sub>3</sub><sup>+</sup>)=250 Hz and  $\Delta v$ (CH<sub>2</sub>)=230 Hz). The <sup>1</sup>H scale was calibrated with external standard – glycine (low-field NH<sub>3</sub> signal at 8.0 ppm and the high field  $\alpha$ –H signal at 2.5 ppm. The total experimental time was ca. 0.5-1 hour.

2D 1H-13C FSLG HETCOR MAS NMR experiments: Two-dimensional (2D)  ${}^{1}\text{H}{}^{-13}\text{C}$ HETCOR experiments were performed using the FSLG (Frequency Switched Lee-Goldburg) decoupling during the  $t_1$  evolution period consisting of 64 increments each made of 128-256 scans with a dwell time of 42.6 µs (**Figure S1**). Rotation frequency was  $\omega_r/2\pi = 11$  kHz. The B<sub>1</sub>( ${}^{1}\text{H}$ ) field strength of FSLG and SPINAL-64 decoupling expressed in frequency units  $\omega_1/2\pi = \gamma B_1$  was 89.3 kHz. The total experimental time was typically 3-5 days.



Figure S1. Schematic representation of 2D <sup>1</sup>H-<sup>13</sup>C FSLG HETCOR MAS NMR experiment.

<sup>*I*</sup>*H*-<sup>*I*</sup>*H SQ/SQ DUMBO NMR:* The 2D <sup>1</sup>H-<sup>1</sup>H SQ/SQ DUMBO NMR correlation spectra were measured using the NOESY-type pulse sequence with DUMBO homo-decoupling applied in both detection periods (**Figure S2**). The recycle delay was 30 s,  $t_1$  evolution period consisted of 128 increments each made of 64-128 scans. The spin-diffusion period (SD) was varied from 20 to 300 µs. The 90° (<sup>1</sup>H) pulse-length was 2.2 µs, power level for DUMBO shape pulse was 71 W, and DUMBO pulse length was 32 µs. All parameters were optimized on glycine to reach maximum spectral resolution ( $\Delta v(NH_3^+)=250$  Hz and  $\Delta v(CH_2)=230$  Hz, **Figure S3**). The <sup>1</sup>H scale was calibrated with external standard – glycine (low-field NH<sub>3</sub> signal at 8.0 ppm and the high field  $\alpha$ -H signal at 2.5 ppm. The total experimental time was typically 3-5 days.



**Figure S2**. Schematic representation of 2D  $^{1}$ H- $^{1}$ H SQ/SQ DUMBO NMR experiment with a spin-diffusion period.



**Figure S3**. 2D  $^{1}$ H-<sub>1</sub>H SQ/SQ DUMBO NMR spectrum (10 kHz) of glycine measured with a 500 µs mixing period.

<sup>1</sup>*H*-<sup>1</sup>*H DQ/SQ DUMBO NMR correlation experiments:* The 2D <sup>1</sup>H-<sup>1</sup>H DQ/SQ DUMBO NMR correlation spectra were measured using the <sup>1</sup>H-<sup>1</sup>H double-quantum (DQ) experiment employing the SPC5 recoupling sequence at spinning frequency  $\omega_t/2\pi = 10$  kHz (**Figure S4**). The recycle delay was 30 s,  $t_1$  evolution period consisted of 128 increments each made of 64-128 scans. The DQ coherence excitation and reconversion consisted of 1-4 loops (duration of one loop was 40 µs). The DUMBO decoupling was applied during both detection periods. Similarly as in the previous case all the experimental parameters were optimized on glycine sample. The total experimental time was typically 3-5 days.



**Figure S4**. Schematic representation of 2D <sup>1</sup>H-<sup>1</sup>H DQ/SQ DUMBO NMR experiment with SPC5recoupling sequence.

2D <sup>1</sup>H-<sup>13</sup>C PILGRIM MAS NMR: The site-specific measurement of one-bond <sup>1</sup>H-<sup>13</sup>C dipolar couplings under the Lee-Goldburg condition was achieved by the 2D PILGRIM experiment (**Figure S5**). The length of polarization-inversion period was 1 ms. Lee-Goldburg cross polarization was incremented from 50 to 5170 µs with 20 µs increment. The experiments were performed at spinning frequency  $\omega_r/2\pi = 10$  kHz. The recycle delay was 30 s,  $t_1$  evolution period consisted of 32 increments each made of 128 scans. Total experimental time was ca. 35 hours. To obtain correct values of <sup>1</sup>H-<sup>13</sup>C dipolar coupling the indirect  $F_1$  axis must be scaled by the factor 0.557=sin(54.7)/ $\sqrt{2}$ .



**Figure S5.** Schematic representation of 2D <sup>1</sup>H-<sup>13</sup>C PILGRIM experiment.

#### Supporting Information SI3 – Computational procedures

The periodic structures of the candidate polymorphs were generated using the Polymorph Predictor module of the Materials Studio Package<sup>10</sup> as described in our previous work.<sup>11</sup> The selected structures were subjected to the full geometrical optimization using the PW DFT approach with the periodic boundary conditions imposed, as implemented in the CASTEP 6.1 suite of codes.<sup>12-14</sup> The resulting geometries served as the input for the NMR chemical shielding predictions, which were performed by applying the GIPAW method<sup>15,16</sup> implemented in the CASTEP-NMR module<sup>13</sup>. The RPBE<sup>17</sup> and PBE<sup>18</sup> functionals were employed in the CASTEP calculations of DAC and SA, respectively. The chemical shielding data were then used to evaluate the agreement between the theoretical and measured values through the protocol described previously.<sup>11</sup> See sections below for further details.

#### Technical details of the calculations

The procedure was followed, which has been successfully employed in the previous investigation of DAC-I (discussed in the main text).<sup>11</sup> Thus, the 'Fine' level of settings corresponding to the CASTEP implementation in Materials Studio 5.0 was kept. In particular, the cut-off energy value of the plane waves was 550 eV, and the same Monkhorst–Pack grids<sup>19</sup> as specified in part S3 of the supplementary information to ref. 11 were used. The CASTEP defaults were used for all the remaining settings, including an application of the on-the-flight generated ultrasoft pseudopotentials.<sup>20</sup>

For sebacic acid, the Brillouin zone in calculations of the polymorphs generated in the space group P2<sub>1</sub>/c was:  $7 \times 1 \times 5$  Monkhorst–Pack grid, 18 *k*-points used; in the space group P1:  $3 \times 7 \times 2$  Monkhorst–Pack grid, 21 *k*-points used; in the space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>:  $3 \times 1 \times 6$  Monkhorst–Pack grid, 6 *k*-points used; in the space group C2/c:  $6 \times 6 \times 1$  Monkhorst–Pack grid, 12 *k*-points used; in the space group P2<sub>1</sub>:  $2 \times 5 \times 4$  Monkhorst–Pack grid, 12 *k*-points used.

#### Data evaluation

The quantification of the level agreement between the theoretical and experimental NMR parameters was performed (without any attempts at the referencing)<sup>21</sup> adopting the approach developed previously<sup>22,23</sup> which is summarized below.

In a typical approach, it is assumed that for the set of q NMR-active nuclei of an isotope Q, the value of the isotropic chemical shift of the *l*-th nucleus from this set,  $\delta(Q)_l$ , was established experimentally, while the value of the isotropic chemical shielding of this nucleus,  $\sigma(Q)_l$ , was obtained from a theoretical calculation; subsequently the isotropic shieldings are fitted, in the least-squares sense, to the measured chemical shifts:

$$\min_{\sigma(\mathbf{Q}); \alpha, \beta} \sum_{l=1}^{q} (\alpha * \sigma(\mathbf{Q})_l + \beta - \delta(\mathbf{Q})_l)^2$$

in order to first obtain the slope,  $\alpha$ , and the intercept,  $\beta$ . The parameters  $\alpha$  and  $\beta$  are then used to calculate the residuals,  $\gamma(Q)_l$ ,

$$\gamma(Q)_l = \alpha * \delta(Q)_l + \beta - \sigma(Q)_l$$

and the mean of these residuals,  $\bar{\gamma}_Q$ ,

$$\bar{\gamma}_{\rm Q} = \frac{1}{q} \sum_{l=1}^{q} \gamma({\rm Q})_l$$

to arrive at the standard deviation of the observations for the Q, SD(Q):

$$SD(\mathbf{Q}) = \sqrt{\frac{1}{q-1} \sum_{l=1}^{q} (\gamma(\mathbf{Q})_l - \bar{\gamma}_{\mathbf{Q}})^2}$$

In order to avoid any confusion, the symbol SD(Q) is used for this standard deviation, instead of  $\sigma$  (which also is a symbol for the NMR chemical shielding) or  $s_Q$  (which is employed for an analogous parameter defined below).

In an alternative approach<sup>22,23</sup> it is assumed that the measured values of the isotropic chemical shifts,  $\delta(X)_i$  and  $\delta(Y)_j$ , with i = 1, ..., n; j = 1, ..., m and  $n \le m$ , of the nuclei X and Y forming the correlation pairs  $[\delta(X)_k; \delta(Y)_k]$  for k = 1, ..., m, can be similarly fitted to the corresponding isotropic chemical shieldings,  $\sigma(X)_i$  and  $\sigma(Y)_j$ , by minimizing

$$\min_{\delta(\mathbf{X}); a, b} \sum_{k=1}^{m} (a * \delta(\mathbf{X})_k + b - \sigma(\mathbf{X})_k)^2$$

and

$$\min_{\delta(\mathbf{Y}); c, d} \sum_{k=1}^{m} (c * \delta(\mathbf{Y})_k + d - \sigma(\mathbf{Y})_k)^2$$

in order to subsequently extract the sets,  $\varepsilon(X)_k$  and  $\varepsilon(Y)_k$ , of so called theoretical chemical shifts,  $\varepsilon$ , from

$$\varepsilon(\mathbf{X})_k = a * \sigma(\mathbf{X})_k + b; \ \varepsilon(\mathbf{Y})_k = c * \sigma(\mathbf{Y})_k + d$$

Within the pair-list of the above-mentioned correlation pairs, the differences between the measured and theoretical chemical shifts are given by

$$\pi(\mathbf{X})_k = \delta(\mathbf{X})_k - \varepsilon(\mathbf{X})_k; \rho(\mathbf{Y})_k = \delta(\mathbf{Y})_k - \varepsilon(\mathbf{Y})_k$$

and the means by

$$\bar{\pi}_{\mathbf{X}} = \frac{1}{m} \sum_{k=1}^{m} \pi(\mathbf{X})_k$$

and

$$\bar{\rho}_{\mathbf{Y}} = \frac{1}{m} \sum_{k=1}^{m} \rho(\mathbf{Y})_k$$

By employing the aforementioned parameters, the similarity of the  $[\delta(X)_k; \delta(Y)_k]$  and  $[\varepsilon(X)_k; \varepsilon(Y)_k]$  sets of the pairs of the chemical shifts can be quantified by the covariance,  $s_{XY}$ , defined as

$$s_{XY} = \frac{1}{m-1} \sum_{k=1}^{m} (\pi(X)_k - \bar{\pi}_X) (\rho(Y)_k - \bar{\rho}_Y)$$

and by the standard deviations

$$s_{\rm X} = \sqrt{\frac{1}{m-1} \sum_{k=1}^{m} (\pi({\rm X})_k - \bar{\pi}_{\rm X})^2}$$

and

$$s_{\rm Y} = \sqrt{\frac{1}{m-1} \sum_{k=1}^{m} (\rho({\rm Y})_k - \bar{\rho}_{\rm Y})^2}$$

## The description of the program POSEL

The current version of POSEL is written in FORTRAN77 and takes the input from up to seven files (in free format), which have to be present in its working directory:

- 1. 'data.txt' (mandatory) is the command file described below;
- 2. 'shiftsC.txt' (mandatory) presumably contains the 13-C NMR chemical shifts of all carbon nuclei of an investigated system; they need to be ordered to comply with the specifications in 'data.txt';
- 3. 'shiftsH.txt' (mandatory) presumably contains the 1-H NMR chemical shifts of all hydrogen nuclei of an investigated system; they need to be ordered to comply with the specifications in 'data.txt';
- 4. 'shiftsN.txt' (optional) presumably contains the 15-N NMR chemical shifts of all nitrogen nuclei of an investigated system;
- 5. 'C.txt' (mandatory) contains the 13-C NMR chemical shielding of the carbon nuclei, ordered as in the corresponding 'shiftsC.txt' file, for the trial structures of an investigated system;
- 6. 'H.txt' (mandatory) contains the 1-H NMR chemical shielding of the hydrogen nuclei, ordered as in the corresponding 'shiftsH.txt' file, for the trial structures of an investigated system;
- 7. 'N.txt' (mandatory if 'shiftsN.txt' is supplied) contains the 15-N NMR chemical shielding of the nitrogen nuclei, ordered as in the corresponding 'shiftsN.txt' file, for the trial structures of an investigated system.

In 'data.txt' file, the items are specified in the following order:

- 1. the threshold value, *SD*(H), of the required level of agreement between the 1-H NMR chemical shieldings and chemical shifts, expressed as the r. m. s. d. of the corresponding fit;
- 2. the threshold value, *SD*(C), of the required level of agreement between the 1-H NMR chemical shieldings and chemical shifts, expressed as the r. m. s. d. of the corresponding fit;
- 3. the number of trial structures;
- 4. the total number of hydrogen nuclei;

- 5. the number of hydrogen nuclei which are considered not to be involved in hydrogen bonds;
- 6. the total number of carbon nuclei;
- 7. the number of carbon nuclei which have two protons attached to them;
- 8. the number of nitrogen nuclei.

In the case of decitabine, the values for items 4–8 would of course be (see **Figure S6**): twelve for #4; eight for #5 (shown in blue and in orange in **Figure S6**); eight for #6; two for #7 (shown in cyan in **Figure S6**); and four for #8. It is stressed that the chemical shifts (and the corresponding chemical shieldings) have to reflect the explicit assignment of the NMR signals of an investigated system.

The output of a POSEL run is provided in the file 'RESULTS.TXT' containing the similarity measures for the structures, which featured a better agreement with experiment than given by both the SD(H) and SD(C) values. Those structures are ordered according to the decreasing value of the covariance.



Figure S6. Pooling the 1-H, 13-C and 15-N nuclei of decitabine.

# Supporting Information SI4 – Long-range <sup>1</sup>H-<sup>1</sup>H correlation spectroscopy



**Figure S7**. <sup>1</sup>H-<sup>1</sup>H CRAMPS spin-diffusion MAS NMR spectrum of DAC/PSA-*co*-PCH microbeads measured with 10 ms spin-diffusion mixing time.

# Supporting Information SI5 – Isotropic values of <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N NMR chemical shifts of Decitabine Polymorphic Form X (DAC-X)

**Table S1.** Experimentally determined isotropic  ${}^{13}C$ ,  ${}^{1}H$  and  ${}^{15}N$  NMR chemical shifts for decitabine polymorphic Form X (DAC-X) found in the synthesized of microbeads formulations.

DAC-X (experimental data)							
<sup>13</sup> C NMR, ppm	1 <sup>1</sup> H	<sup>1</sup> H NMR, ppm		<sup>15</sup> N NMR, ppm			
C1' 91.06	H1'	7.23	N1	175.09			
C2' 42.39	H2'*	1.25	N3	216.79			
	H2'**	2.40	N5	194.71			
C3' 73.96	H3'	3.71	$NH_2$	95.94			
C4' 89.92	H4'	3.92					
C5' 64.32	H5'*	2.89					
	H5'**	3.31					
C2 157.50	)						
C4 167.93	3						
C6 154.96	6 H6	8.39					
	NH <sub>2</sub> *	7.25					
	NH <sub>2</sub> **	7.81					
	ОН	4.50 or 5.00					
	ОН	4.50 or 5.00					



**Figure S8**. Numbering of <sup>1</sup>H, <sup>13</sup>C atoms of decitabine.

# Supporting Information SI6 – Similarity measures for DAC-I (reference) and DAC-X

**Table S2.** The statistical parameters describing the level of agreement between the GIPAW-RPBE predicted and experimental NMR data: *i*) the first column describes agreement between the best predicted structure CSP#A04 and the experimental data obtained for the reference polymorphic form DAC-I; *ii*) the second column represents agreement between the XRPD refined structure of the reference polymorphic form DAC-I and the experimental NMR data; and *iii*) the third column demonstrates agreement between the best predicted structure CSP#25 and the experimental data obtained for the unknown polymorphic form DAC-X.

similarity measure	(DFT-calculated)/(experimental) NMR parameters			
	CSP#A04/DAC-I	XRPD/DAC-I	CSP#25/DAC-X	
<sup>13</sup> C (all) r.m.s.d, ppm	2.34	2.68	1.92	
<sup>1</sup> H (all) r.m.s.d., ppm	0.378	0.376	0.428	
<sup>1</sup> H (CH) r.m.s.d., ppm	0.347	0.203	0.367	
<sup>1</sup> H (OH/NH) r.m.s.d., ppm	0.145	0.319	0.155	
<sup>15</sup> N r.m.s.d., ppm	2.59	2.11	4.20	
s(CH) covariance, ppm <sup>2</sup>	0.475	0.394	0.209	



The theoretical chemical shifts,  $\varepsilon$ , were generated by first fitting the experimental chemical shifts,  $\delta$ , to the GIPAW-RPBE chemical shieldings,  $\sigma$ :

 $\delta$  = -1.0258\* $\sigma$  + 171.8840 ppm (*i* = 8 data points;  $R^2$  = 0.99695), and subsequently evaluating  $\varepsilon(i)$  at  $\sigma(i)$ .





The theoretical chemical shifts,  $\epsilon$ , were generated by first fitting the experimental chemical shifts,  $\delta$ , to the GIPAW-RPBE chemical shieldings,  $\sigma$ :

δ = 1.0224\*σ + 170.8288 ppm (i = 8 data points;  $R^2$  = 0.99804), and subsequently evaluating ε(*i*) at σ(*i*).



**Figure S9.** A graphical representation of the level of agreement between theory and experiment for the best predictions of the two decitabine polymorphs (in a case of a complete agreement, the points would lie on the red line. The experimental NMR parameters of the reference system DAC-I compared with the data calculated for best predicted structure CSP#A04 are shown in the left column, whereas the experimental NMR parameters of the system DAC-X compared with the data calculated for best predicted structure CSP#25 are demonstrated in the right column.



For all  $^{13}\text{C}$  nuclei, the GIPAW-RPBE chemical shieldings,  $\sigma$ , were fitted to the experimental chemical shifts:

 $\sigma$  = -0.9764\* $\delta$  + 166.9806 ppm (8 data points;  $R^2$  = 0.99804)



For all  ${}^{1}$ H nuclei, the GIPAW-RPBE chemical shieldings,  $\sigma$ , were fitted to the experimental chemical shifts:

 $\sigma$  = -0.8056\* $\delta$  + 29.6800 ppm (12 data points;  $R^2$  = 0.94606)



For all  $^{15}\text{N}$  nuclei, the GIPAW-RPBE chemical shieldings,  $\sigma$ , were fitted to the experimental chemical shifts:

 $\sigma = -1.0457^*\delta + 227.6036 \text{ ppm}$ (4 data points;  $R^2 = 0.99131$ )



For the OH/NH  $^1\text{H}$  nuclei, the GIPAW-RPBE chemical shieldings,  $\sigma,$  were fitted to the experimental chemical shifts:

 $\sigma = -1.1295^*\delta + 31.6726 \text{ ppm}$ (4 data points;  $R^2 = 0.98956$ )

**Figure S10.** An evaluation of the level of agreement between theory and experiment for the NMR data of decitabine DAC-X: the best predicted structure CSP#25 was used

# REFERENCES

(1) Sakellariou, D.; Lesage, A.; Hodgkinson, P.; Emsley, L. Homonuclear dipolar decoupling in solid-state NMR using continu-ous phase modulation. *Chem. Phys. Lett.* **2000**, *319*, 253-260.

(2) Schaefer, J.; Stejskal, E. O. Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. *J. Am. Chem. Soc.* **1976**, *98*, 1031-1032.

(3) Wu, X. L.; Burns, S. T.; Zilm, K. W. Spectral Editing in CPMAS NMR. Generating Subspectra Based on Proton Multiplicities. *J. Magn. Reson. Ser. A* **1994**, *111*, 29-36.

(4) vanRossum, B. J.; Forster, H.; deGroot, H. J. M. High-Field and High-Speed CP-MAS13C NMR Heteronuclear Dipolar-Correlation Spectroscopy of Solids with Frequency-Switched Lee–Goldburg Homonuclear Decoupling. *J Magn. Reson.* **1997**, *124*, 516-519.

(5) Sakellariou, D.; Lesage, A.; Emsley, L. Proton–Proton Constraints in Powdered Solids from  ${}^{1}\text{H}{-}^{1}\text{H}{-}^{1}\text{H}$  and  ${}^{1}\text{H}{-}^{1}\text{H}{-}^{13}\text{C}$  Three-Dimensional NMR Chemical Shift Correlation Spectroscopy. *J. Am. Chem. Soc.* **2001**, *123*, 5604-5605.

(6) Brown, S. P.; Lesage, A.; Elena, B.; Emsley, L. Probing Proton–Proton Proximities in the Solid State: High-Resolution Two-Dimensional <sup>1</sup>H–<sup>1</sup>H Double-Quantum CRAMPS NMR Spectroscopy. *J. Am. Chem. Soc.* **2004**, *126*, 13230-13231.

(7) Hohwy, M.; Rienstra, C. M.; Jaroniec, C. P.; Griffin, R. G. Fivefold symmetric homonuclear dipolar recoupling in rotating solids: Application to double quantum spectroscopy. *J. Chem. Phys.* **1999**, *110*, 7983-7992.

(8) Langer, B.; Schnell, L.; Spiess, H. W.; Grimmer, A. R. Temperature Calibration under Ultrafast MAS Conditions. *J. Magn. Reson.* **1999**, *138*, 182-186.

(9) Brus, J. Heating of samples induced by fast magic-angle spinning. *Solid State Nucl. Magn. Reson.* **2000**, *16*, 151-160.

(10) Material Studio 4.4, www.accelrys.com

(11) Brus, J.; Czernek, J.; Kobera, L.; Urbanová, M.; Abbrent, S.; Husak, M. Predicting the Crystal Structure of Decitabine by Powder NMR Crystallography: Influence of Long-Range Molecular Packing Symmetry on NMR Parameters. *Cryst. Growth Des.* **2016**, *16*, 7102–7111.

(12) Segall, M. D.; Lindan, P. J. D.; Probert, M. J.; Pickard, C. J.; Hasnip, P. J.; Clark, S. J.; Payne, M. C. First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys-Condens Mat.* **2002**, *14*, 2717-2744.

(13) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C. First principles methods using CASTEP. *Z Kristallogr.* **2005**, *220*, 567-570.

(14) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, 1758-1775.

(15) Pickard, C. J.; Mauri, F. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B* **2001**, *63*, 245101.

(16) Yates, J. R.; Pickard, C. J.; Mauri, F. Calculation of NMR chemical shifts for extended systems using ultrasoft pseudopoten-tials. *Phys. Rev. B* **2007**, *76*, 024401.

(17) Hammer, B.; Hansen, L. B.; Norskov, J. K. Improved adsorption energetics within density-functional theory using revised Perdew-Burke-Ernzerhof functionals. *Phys. Rev. B* **1999**, *59*, 7413-7421.

(18) Perdew, J. P.; Burke, K., Ernzerhof, M. Generalized Gradi-ent Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.

(19) H. J. Monkhorst, J. D. Pack *Phys. Rev. B* **1976**, *13*, 5188–5192.

(20) P. J. Hasnip, C. J. Pickard Comp. Phys. Commun. 2006, 174, 24–29.

(21) R. K. Harris, P. Hodgkinson, C. J. Pickard, J. R. Yates, V. Zorin, Magn. Reson. Chem. **2007**, *45*, S174–S186.

(22) J. Czernek, J. Brus Chem. Phys. Lett. 2013, 586, 56–60.

(23) J. Czernek, J. Brus Chem. Phys. Lett. 2014, 608, 334–339.