

## Elements of fractal geometry in the $^1\text{H}$ NMR spectrum of a copolymer intercalation-complex: Identification of the underlying Cantor set

John S. Shaw, Rajendran Vaiyapuri, Matthew P. Parker, Claire A. Murray, Kate J. C. Lim, Cong Pan, Marcus Knappert, Christine J. Cardin, Barnaby W. Greenland, Ricardo Grau-Crespo and Howard M. Colquhoun\*

Department of Chemistry, University of Reading, Whiteknights, Reading, RG6 6AD, UK

### SUPPORTING INFORMATION

<u>Information</u>	<u>Page</u>
1. Synthetic, instrumental, analytical and computational methods.	S2
2. Solvent-dependence of $^1\text{H}$ NMR spectra of copolymer <b>1</b> .	S2
3. Computational modelling of pyrene intercalation into copolymer <b>1</b> .	S3
4. Crystal structure analysis of complex <b>3</b> .	S4
5. Methodologies for calculating shielding factors ( <i>T</i> ) and sequence-probabilities ( <i>P</i> ).	S5
6. Simulated $^1\text{H}$ NMR spectra (400 MHz) for copolymer <b>1</b> in the presence of one molar equivalent (relative to diimide units) of pyrene- <i>d</i> <sub>10</sub> , at 400 MHz, from the data for <b>septet</b> sequences in Table 2.	S7
7. Table S1. Total shielding factors and sequence-probabilities for the allowed <b>nonet</b> sequences in copolymer <b>1</b> .	S8
8. Simulated $^1\text{H}$ NMR spectrum (400 MHz) for copolymer <b>1</b> in the presence of one molar equivalent (relative to diimide units) of pyrene- <i>d</i> <sub>10</sub> , from the data for <b>nonet</b> sequences in Table S2.	S9
9. Mathematical formulation of generalized Cantor sets: Demonstration of the quaternary expression for the fourth-quarter Cantor set and generalization to non-integer bases.	S10
E1. Electronic data files for the X-ray structure of complex <b>3</b> [Complex_3.cif; Checkcif_complex_3.pdf; Complex_3_SF.fcf].	
E2. Electronic data file for the energy-minimised structure of the pyrene-intercalated, chain-folded model shown in Figure 3a [Fig_3_model_data.pdb].	

## 1. Synthetic, instrumental, analytical and computational methods.

Synthetic procedures were performed under an atmosphere of dry nitrogen unless otherwise specified. Solvents and reagents were obtained from Sigma-Aldrich and used without purification unless otherwise stated. *N,N*-Dimethylacetamide (DMAc) and *N,N*-dimethyl-formamide (DMF) were distilled over calcium hydride before use. 4,4'-Bis[4-(3-aminophenoxy)benzenesulfonyl]-1,1'-biphenyl was prepared according to a literature procedure,<sup>S1</sup> as was the model oligomer **2**.<sup>S2</sup> Deuterated solvents were purchased from Cambridge Isotope Laboratories. Proton and <sup>13</sup>C NMR spectra were recorded on a Bruker Avance 400 MHz spectrometer, with chemical shifts referenced to residual solvent resonances and reported in ppm relative to TMS. Glass transition temperatures were measured under nitrogen by differential scanning calorimetry (DSC) using a TA Q2000 system, at a heating rate of 10 °C min<sup>-1</sup>. Inherent viscosities ( $\eta_{inh}$ ) were measured at 25 °C in a thermostatted water bath on 0.1% polymer solutions in 1-methylpyrrolidinone (NMP) using a Schott-Geräte CT-52 semi-automated viscometer and AVS 470 measurement system. Gel permeation chromatography was carried out using a Polymer Laboratories PL-220 instrument fitted with 2 x PL 10  $\mu$ m mixed B columns. Molecular weights are referenced to polystyrene standards. Computational modelling using molecular mechanics with charge-equilibration, (DREIDING force-field, *Materials Studio*, v. 7.0, Accelrys Inc., San Diego) was carried out in order to obtain a preliminary structural model for copolymer-pyrene interactions. Atomic coordinates for the modelled structure (main paper Figure 3a) are provided in electronic format as a PDB file. Simulated NMR spectra were generated using the "Peak Table to Spectrum" script in Mestrenova (v. 9.0, Mestrelab Research, Santiago de Compostela). Single crystal X-ray data were measured for the pyrene complex **3** on an Oxford Diffraction X-Calibur Gemini diffractometer using Cu-K $\alpha$  radiation at 150 K. Details of structure solution and refinement are given below and, in electronic format, in the associated cif, Checkcif, and structure-factor files.

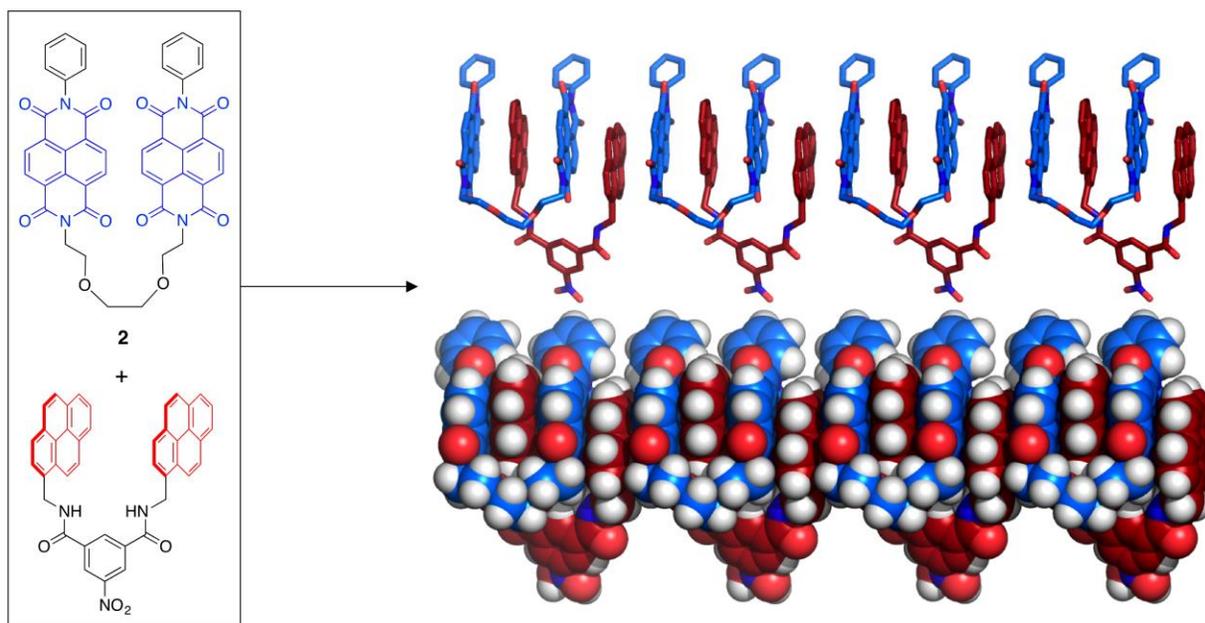
## 2. Solvent-dependence of <sup>1</sup>H NMR spectra of copolymer **1** in the diimide region

It might be expected that the resonances associated with inequivalent pairs of protons in unsymmetrically-substituted naphthalenediimide residues (Al-NDI-Ar) would show spin-spin splitting, and this is indeed evident when CDCl<sub>3</sub>/CF<sub>3</sub>COOH (6:1 v/v) is used as solvent for copolymer **1**. However, the coupling is suppressed when CDCl<sub>3</sub>/(CF<sub>3</sub>)<sub>2</sub>CHOH (6:1 v/v) is used,

and the latter solvent mixture was employed exclusively in the present complexation study. Spin-spin splitting of coupled, inequivalent protons vanishes when the two components of the spin system, despite their inequivalence, have fortuitously coincident chemical shifts.<sup>S4</sup> This is evidently the case for copolymer **1** and its complexes with pyrene when CDCl<sub>3</sub>/(CF<sub>3</sub>)<sub>2</sub>CHOH (6:1, v/v) is used as <sup>1</sup>H NMR solvent.

### 3. Computational modelling of pyrene intercalation into copolymer **1**.

The starting point for modelling a pyrene complex of the chain-folded copolymer **1** was the recently-reported X-ray structure of a complex between the bis-diimide **2** and a bispyrenyl tweezer-molecule (Figure S1).<sup>S2</sup>



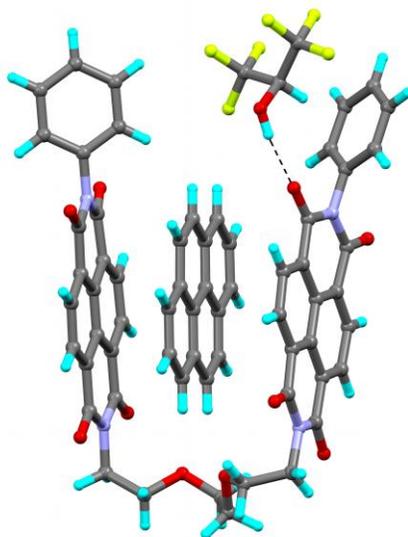
**Supporting Figure S1.** Formation and X-ray crystal structure of a complex between bis-diimide **2** and a bispyrenyl tweezer-molecule.<sup>S2</sup>

The segment of copolymer **1** seen in Figure 3a (main paper) was generated in *Materials Studio* by first deleting the *N*-phenyl substituents from the molecules of **2** in the structure shown above, and replacing them with linking triethylenedioxy units and a terminal phenylsulfonyl(4,4'-biphenylene)sulfonyl4-phenoxyphenyl substituent. The bis(pyrenyl) tweezer molecules were converted to simple pyrene molecules by deleting the linking aromatic diamide groups and

replacing them with hydrogens. The resulting model was then energy-minimised in *Materials Studio* (molecular mechanics with charge-equilibration) using the DREIDING force-field.<sup>S5</sup> Atomic coordinates for the final model are given in the SI file [Fig\_3\_model\_data.pdb].

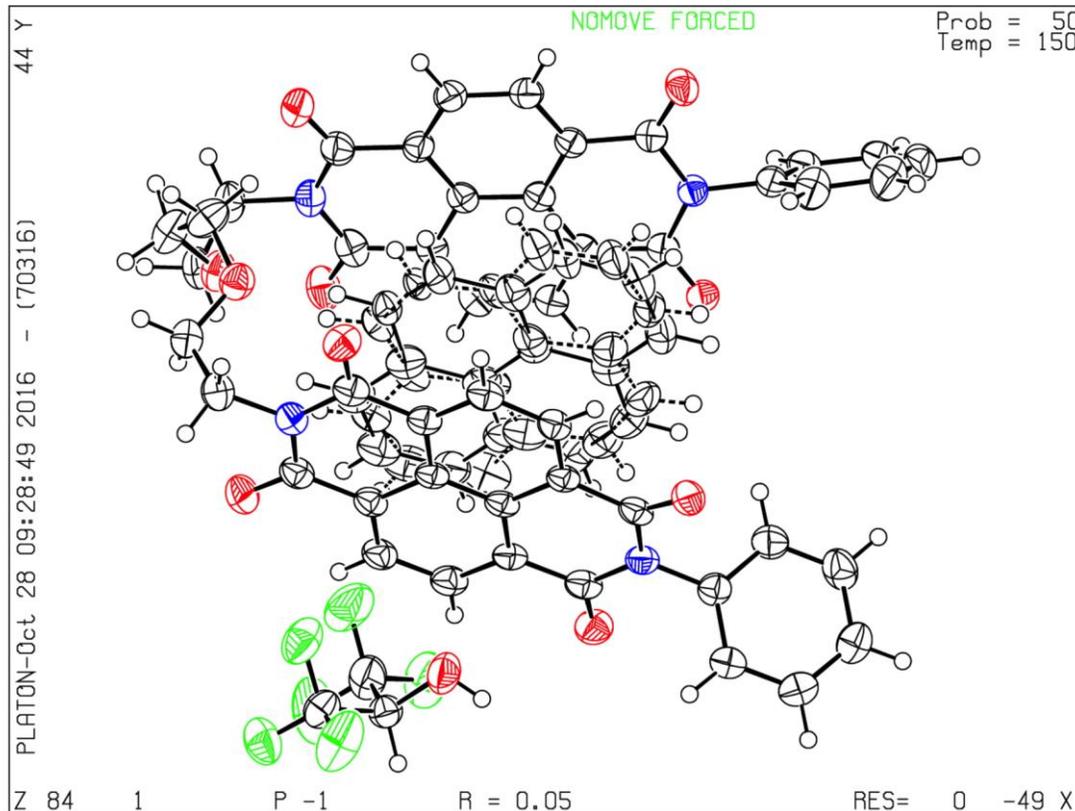
#### 4. Crystal structure analysis of complex **3**

Dark red single crystals of complex **3** were grown from an equimolar solution of compound **2** and pyrene in chloroform/hexafluoropropan-2-ol (6:1 v/v) by vapour diffusion against methanol. Crystal data for **3**: C<sub>46</sub>H<sub>30</sub>N<sub>4</sub>O<sub>10</sub>.C<sub>16</sub>H<sub>10</sub>.(CF<sub>3</sub>)<sub>2</sub>CHOH, *M*<sub>r</sub> = 1169.06, triclinic, *P*-1, *a* = 10.6269(3), *b* = 14.9210(5), *c* = 17.4851(7) Å,  $\alpha$  = 108.065(3),  $\beta$  = 102.085(3),  $\gamma$  = 91.766(2)°. *V* = 2563.73(11) Å<sup>3</sup>, *T* = 150(2) K, *Z* = 2, *D*<sub>c</sub> = 1.515 g cm<sup>-3</sup>,  $\mu$ (Cu-K $\alpha$ ) = 1.006 mm<sup>-1</sup>, *F*(000) = 1204. Independent measured reflections 8183. *R*<sub>1</sub> = 0.0516, *wR*<sub>2</sub> = 0.0927 for 6639 independent observed reflections [ $2\theta \leq 126^\circ$ ,  $I > 2\sigma(I)$ ]. CCDC 1503298. The crystal contains one molecule of hexafluoropropan-2-ol per supramolecule of **3**, hydrogen bonded to a diimide carbonyl oxygen as shown below. The bound pyrene molecule was disordered over two sites, and was refined using parts with the total occupancy set to 1.



**Supporting Figure S2.** X-ray structure of complex **3**, showing the solvating molecule of hexafluoropropan-ol [hydrogen bond O.....O distance = 2.692(1)Å]. Only one of the two sites of pyrene intercalation is shown. The second site lies in the same plane as the first and is related to it by a ca. 10° rotation. See Supporting Figure S3 and the corresponding cif.

Datablock 1 - ellipsoid plot



**Supporting Figure S3.** X-ray structure of complex **3**, showing thermal ellipsoids. The pyrene guest molecule is disordered over two sites, as shown.

## 5. Methodologies for calculating shielding factors ( $T$ ) and sequence-probabilities ( $P$ ).

Total shielding factors,  $T$ , were obtained for any given septet sequence by determining its three-digit code (main paper, Table 1) from the numbers and positions of "II" pairs in the sequence. The three digits were then used as coefficients  $N_k$  in the summation:

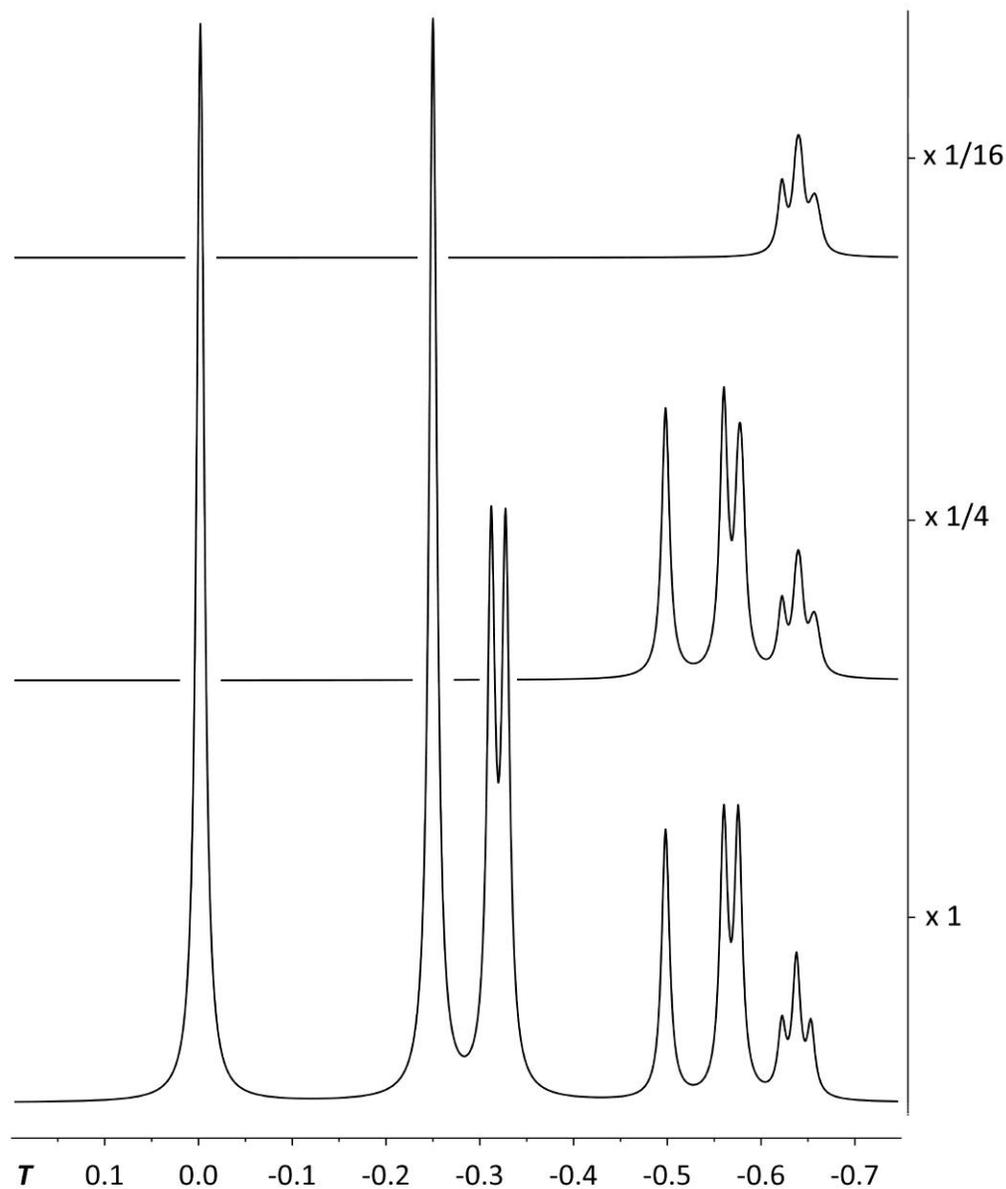
$$T = \sum_{k=1}^3 \frac{N_k}{4^k} \quad (\text{Equation 1})$$

to give the total shielding factor for that sequence. It should be noted that the starting value of  $k$  (i.e. 1) is not fixed by the model but is in fact a completely arbitrary integer. Absolute values of  $T$

are thus scalable, to any magnitude, by powers of 4. The *pattern* of shielding factors however remains unchanged whatever the starting value of  $k$ , a classic example of the scale-independence of self-similar structures. Our choice of 1 as the initial value was the obvious place to start, but it also proved to have the additional advantage that the range of  $T$ -values so obtained (0 to 0.65625; Table S1) then corresponds very closely to the range of observed complexation shifts ( $\Delta\delta$ ) in ppm (main paper, Figure 1). The terminating value of  $k$  (here 3) is determined by the length of the sequence being considered. Thus, for septets, there are 3 possible positions for "II" pairs relative to the central "observed" diimide residue. For nonets there are 4 such positions; for undecets 5, and so on. The limiting value of  $T$ , for the infinite sequence IIII..... is  $2/3$ , corresponding to  $k = 1$  to  $\infty$  for  $Nk$  always = 2.

**The probability** of any given "I"-centred sequence was calculated by assigning an *a priori* probability of one to the central "I" residue, and then – working outwards from this residue in both directions – multiplying the probabilities of finding the next-adjacent residues. For example in random copolymer **1**, which contains equal numbers of "I" and "S" residues, the septet sequence IIS**I**SII has the probability  $0.5 \times 1 \times 0.5 \times 1 \times 0.5 \times 1 \times 0.5 = 0.0625$ . Individual residue-probabilities are assigned from the chemistry-based rule (see main paper) that "I" can be followed with equal probability by either "I" or "S", but "S" can be followed only by "I". The sequence IIS**I**SII is symmetrical, but many other sequences, e.g. IIS**III** are unsymmetrical and so – in NMR terms – are indistinguishable from their mirror images, in this case III**I**SII. We term such unsymmetrical sequences *directionally degenerate*. As these comprise two equally probable but spectroscopically indistinguishable sequences, they are represented in Tables S1 and S2 by a single sequence with twice the probability of a symmetrical one. Again, the method is not limited to septets. Sequence probabilities (and shielding factors) for **nonets** are thus enumerated as a further example of the method in Table S2, and are used to generate the simulated NMR spectra shown in Figure S3. The self-consistency of the method is confirmed by the fact that the probabilities for all possible sequences of a given length [final columns of Table 2 (main paper) and Table S1] sum precisely to 1.

6. Simulated  $^1\text{H}$  NMR spectra (400 MHz) for copolymer **1** in the presence of one molar equivalent (relative to diimide units) of pyrene- $d_{10}$ , at 400 MHz. These spectra were simulated using the data for septet sequences in Table 2 (main paper) at different length scales but at a constant, realistic, spectroscopic linewidth (4 Hz). The shielding factor axis  $T$  refers to the original, unscaled ( $\times 1$ ) spectrum

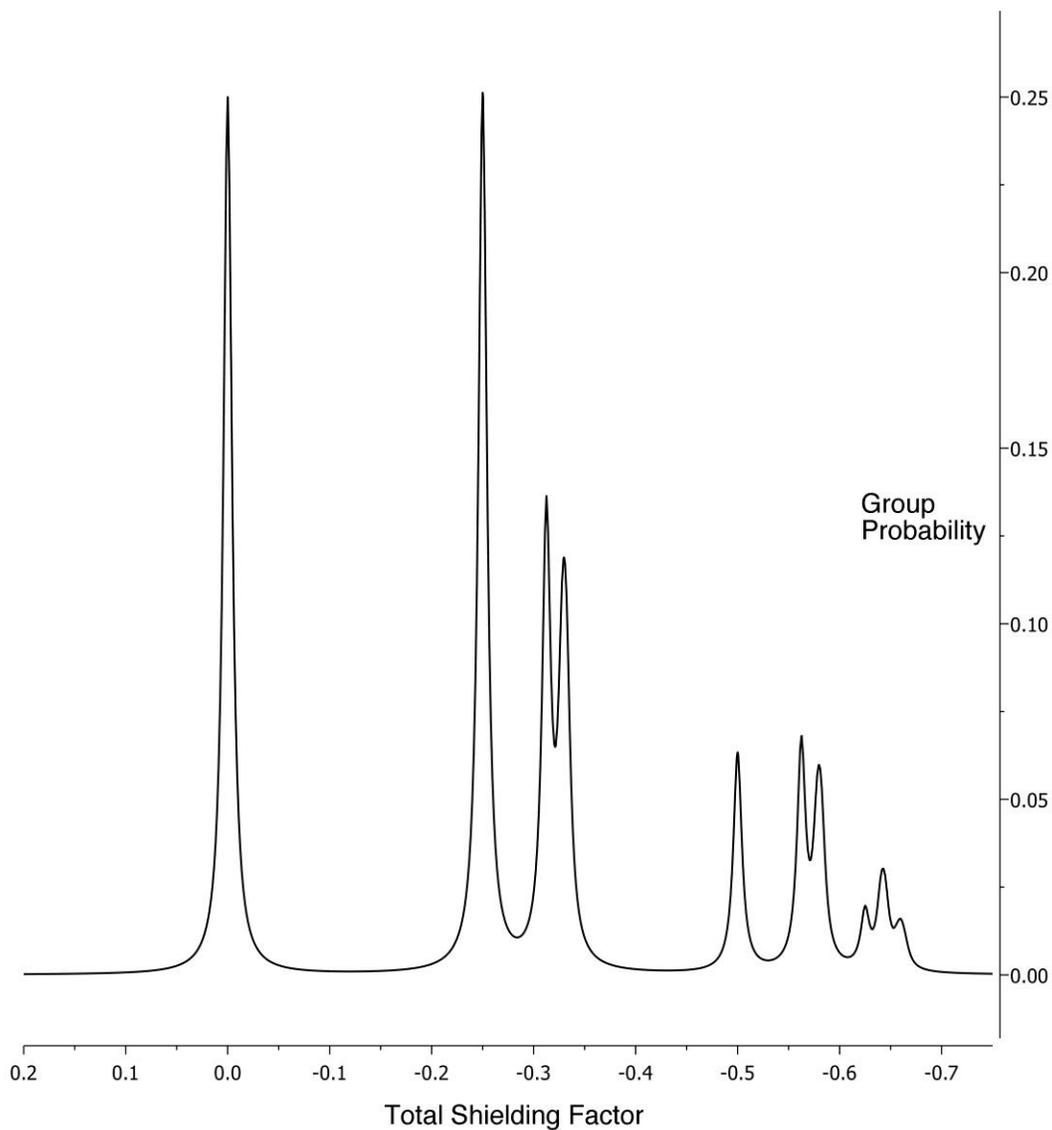


**Supporting Figure S4.** The re-scaled spectra are aligned to highlight the self-similar character of the " $\times 1$ " spectrum. The unscaled spectrum, based on the septet sequences in Table S1, uses  $T$ -values from the fourth-quarter Cantor set (Table S1) as complexation shifts ( $\Delta\delta$ ), and sequence probabilities as integrals.

**7. Table S1.** Total shielding factors and sequence-probabilities for the allowed **nonet** sequences in copolymer 1. The simulated <sup>1</sup>H NMR spectrum shown in Figure S4 was again generated by plotting "Group probability" against "Group T-value".

Nonet sequence	Digital code	Shielding factor $T (= \Delta\delta)$	Group T-value (= $\Delta\delta$ )	Sequence probability (P)	Directional degeneracy (D)	Overall probability (P × D)	Group probability [ $\Sigma(P \times D)$ ]
ISISISII	0000	0		0.0625	1	0.0625	
ISISISII	0000	0		0.03125	2	0.0625	
ISISISIII	0000	0		0.03125	2	0.0625	
SIISISII	0000	0		0.015625	1	0.015625	
SIISISIII	0000	0		0.015625	2	0.03125	
IIISISIII	0000	0	<b>0</b>	0.015625	1	0.015625	<b>0.25</b>
SISIIISII	1000	0.25		0.03125	2	0.0625	
SISIIISII	1000	0.25		0.03125	2	0.0625	
SISIIISII	1000	0.25		0.015625	2	0.03125	
SISIIISIII	1000	0.25		0.015625	2	0.03125	
IIISIIISII	1000	0.25	<b>0.25</b>	0.015625	2	0.03125	<b>0.25</b>
IIISIIISIII	1000	0.25		0.015625	2	0.03125	
ISIIISISII	1100	0.3125		0.03125	2	0.0625	
ISIIISISII	1100	0.3125		0.015625	2	0.03125	
ISIIISISIII	1100	0.3125	<b>0.3125</b>	0.015625	2	0.03125	<b>0.125</b>
SIIIIISISII	1110	0.328125		0.015625	2	0.03125	
SIIIIISISII	1110	0.328125		0.0078125	2	0.015625	
SIIIIISISIII	1110	0.328125	<b>0.328125</b>	0.0078125	2	0.015625	<b>0.0625</b>
IIIIISISII	1111	0.33203125		0.015625	2	0.03125	
IIIIISISII	1111	0.33203125		0.0078125	2	0.015625	
IIIIISISIII	1111	0.33203125	<b>0.33203125</b>	0.0078125	2	0.015625	<b>0.0625</b>
SISIIISISII	2000	0.5		0.015625	1	0.015625	
SISIIISISII	2000	0.5		0.015625	1	0.015625	
IIISIIISISII	2000	0.5		0.015625	1	0.015625	
IIISIIISISII	2000	0.5	<b>0.5</b>	0.015625	1	0.015625	<b>0.0625</b>
SISIIISISII	2100	0.5625		0.0156125	2	0.03125	
IIISIIISISII	2100	0.5625	<b>0.5625</b>	0.0156125	2	0.03125	<b>0.0625</b>
SISIIISISII	2110	0.578125		0.0078125	2	0.015625	
IIISIIISISII	2110	0.578125	<b>0.578125</b>	0.0078125	2	0.015625	<b>0.03125</b>
SISIIISISII	2111	0.58203125		0.0078125	2	0.015625	
IIISIIISISII	2111	0.58203125	<b>0.58203125</b>	0.0078125	2	0.015625	<b>0.03125</b>
ISIIIIISII	2200	0.625	<b>0.625</b>	0.015625	1	0.015625	<b>0.015625</b>
ISIIIIISII	2210	0.640625	<b>0.640625</b>	0.0078125	2	0.015625	<b>0.015625</b>
ISIIIIISII	2211	0.64453125	<b>0.64453125</b>	0.0078125	2	0.015625	<b>0.015625</b>
SIIIIISISII	2220	0.65625	<b>0.65625</b>	0.00390625	1	0.00390625	<b>0.00390625</b>
SIIIIISISII	2221	0.66015625	<b>0.66015625</b>	0.00390625	2	0.00390625	<b>0.0078125</b>
IIIIIIISII	2222	0.6640625	<b>0.6640625</b>	0.00390625	1	0.00390625	<b>0.00390625</b>

8. Simulated  $^1\text{H}$  NMR spectrum (400 MHz) for copolymer **1** in the presence of one molar equivalent (relative to diimide units) of pyrene- $d_{10}$ , from the data for **nonet** sequences shown in Table S1 above.



**Supporting Figure S5.** This nonet-based simulation (400 MHz) was generated from the 15 values of  $T$  given in Table S2, but shows only the same 10 resonances as the septet-based Supporting Figure S2 because of peak overlap at the realistic linewidth (4 Hz) used in the simulation. The resulting match to experiment (main paper Figure 2) is rather better for nonets than for septets, notably in terms of the asymmetry of several resonance-groups, but the septet-based simulation remains a good approximation.

## 9. Mathematical formulation of generalized Cantor sets

### 9.1 Demonstration of the quaternary expression for the fourth-quarter Cantor set

The fourth-quarter Cantor set, like the middle-third Cantor set, can be defined as the limit of a sequence of sets  $C_k$ . We start with the closed interval  $C_0 = [0,1]$ . Then the set  $C_1$  is obtained by removing the fourth quarter (of length  $\frac{1}{4}$ ) from  $C_0$ , leaving  $[0, \frac{3}{4}]$ . The next set  $C_2$  is defined by removing the fourth quarter (of length  $\frac{1}{16}$ ) from each of the three remaining quarters of length  $\frac{1}{4}$ . So  $C_2 = [0, \frac{3}{16}] \cup [\frac{1}{4}, \frac{7}{16}] \cup [\frac{1}{2}, \frac{15}{16}]$ . The set  $C_3$  is defined by removing the fourth quarter (of length  $\frac{1}{64}$ ) from each of the remaining quarters of length  $\frac{1}{16}$ . And so on *ad infinitum*. The fourth-quarter Cantor set is the limit of the sequence  $C_k$  when  $k$  goes to infinity. Since  $C_0 \supseteq C_1 \supseteq C_2 \supseteq \dots$ , then we can also define the limit to be the intersection of the sets:

$$C = \bigcap_{k=1}^{\infty} C_k \quad \text{Eq. S1}$$

Mandelbrot coined the word “trema”, from a Latin word meaning “hole”, to name the intervals that are removed in this kind of construction.<sup>S6</sup> The trema removed in the first iteration is called the first-level trema; the tremas removed in the second iteration are called second-level tremas, and so on. We demonstrate here the equivalence between the construction by tremas of the fourth-quarter Cantor set and its quaternary expression. The proof is adapted from the one given by Edgar for the middle-third Cantor set.<sup>S7</sup>

**Proposition.** *Let  $x \in [0,1]$ . Then  $x$  belongs to the fourth-quarter Cantor set if and only if  $x$  has a quaternary (base-4) expansion, given by*

$$x = \sum_{k=1}^{\infty} \frac{N_k}{4^k} \quad \text{Eq. S2}$$

*using only the digits  $N_k = 0, 1$  and  $2$ .*

**Proof.** The first place to the right of the point in the quaternary expression of  $x$  is a 3 if and only if  $x$  is between

$$\frac{3}{4} = (0.3)_4 \quad \text{and} \quad 1 = (0.3333\dots)_4 \equiv (0.\bar{3})_4,$$

where we have used the subscript 4 to indicate a number expressed in a quaternary base (instead of in the standard decimal base). The first-level trema is the closed interval  $\left[ (0.3)_4, (0.\bar{3})_4 \right]$ . Neither  $\frac{3}{4}$  nor 1 have expansions excluding the digit 3, so they are both contained in the trema. After this trema is removed from  $C_0$ , we have  $C_1$ . So  $C_1$  contains exactly the numbers in  $[0,1]$  that have a base-4 expansion not using 3 in the first place to the right of the point.

The second place in a number  $x$  in  $C_1$  is a 3 if and only if  $x$  belongs to one of the three second-level tremas:

$$\begin{aligned} \left[ \frac{3}{16}, \frac{1}{4} \right] &= \left[ (0.03)_4, (0.0\bar{3})_4 \right), \\ \left[ \frac{7}{16}, \frac{1}{2} \right] &= \left[ (0.13)_4, (0.1\bar{3})_4 \right), \\ \left[ \frac{11}{16}, \frac{3}{4} \right] &= \left[ (0.23)_4, (0.2\bar{3})_4 \right]. \end{aligned}$$

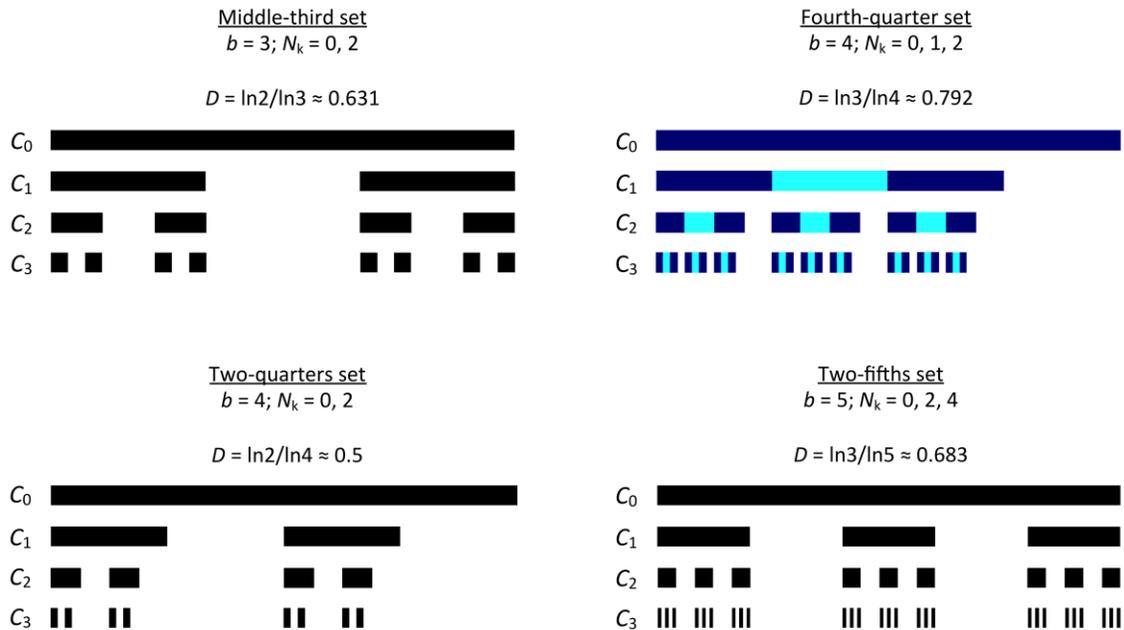
There is a subtlety in whether the upper limit of the interval defining each trema is included in the trema or not. The numbers  $\frac{1}{4}$  and  $\frac{1}{2}$  both have alternative quaternary expressions excluding the digit 3, which are  $\frac{1}{4} = (0.1)_4$  and  $\frac{1}{2} = (0.2)_4$ , so they are not included in the tremas. In contrast, for  $\frac{3}{4}$  the alternative expression  $\frac{3}{4} = (0.3)_4$  still contains a 3, so it must be included in the trema.  $C_2$  is obtained from  $C_1$  after these tremas are removed, so  $C_2$  contains exactly the numbers in  $[0,1]$  that have a base-4 expansion using 3 neither in the first place nor the second place to the right of the point.

In the same way, it is easy to check that  $C_3$  contains exactly the numbers in  $[0,1]$  that have a base-4 expansion not using 3 in the first 3 places after the point. Continuing in this way, we see that the points remaining in the fourth-quarter Cantor set  $C = \bigcap_{k=1}^{\infty} C_k$  are exactly the numbers in  $[0,1]$  that have a base-4 expansion not using 3 at all.

It is possible to generalize the construction rules for Cantor sets based on the expansion formula:

$$x = \sum_{k=1}^{\infty} \frac{N_k}{b^k} \quad \text{Eq. S3}$$

where  $b$  is a positive integer number (greater than 2), and the  $\{N_k\}$  coefficients can take values between 0 and  $b-1$ , except for some excluded values. We start by dividing  $C_0 = [0,1]$  into  $b$  intervals, and assign integers 0 to  $b-1$  to each interval. Then the set  $C_1$  is obtained by removing the intervals corresponding to the excluded integers in expansion S3. Each remaining interval is further divided into  $b$  intervals, which are enumerated 0 to  $b-1$  again. The set  $C_2$  is obtained by removing the open intervals corresponding to the excluded integers. And so on *ad infinitum*.



**Supporting Figure S6.** Some Cantor set constructions (shown up to the third iteration) and their fractal dimensions. The fourth-quarter construction is shown in two colours, because the segments remaining after each division and deletion are contiguous, and the different colours are needed to enable the segments to be visualised. In the other three constructions, the "remaining segments" are non-contiguous and so can be visualised in monochrome.

Fig S5 shows some examples, including the "middle-third" and "fourth-quarter" sets, but also two examples with more than one excluded integer in the expansion. These geometric constructions involve division of a line of unit length into segments, deletion of one or more segments, and then repeating the same two operations on each of the remaining segments. The pattern of the remaining segments is shown *after* each iteration of *both* operations.

The fractal (Hausdorff) dimension of each of these self-similar sets equals their similarity dimension:<sup>1</sup>

$$D = \frac{\ln(\text{Number of allowed integers in } \{N_k\})}{\ln b} \quad \text{Eq. S4}$$

Thus, the fractal dimension of the middle-third set is  $D = \ln 2 / \ln 3 \approx 0.631$ , whereas for the fourth-quarter set it is  $D = \ln 3 / \ln 4 \approx 0.792$ . The limiting case of a non-fractal set occurs when all integers are allowed in the expansion, in which case  $x$  can take all values between 0 and 1, and then  $D = 1$ .

### 9. 2 Generalization of the Cantor set to non-integer bases

We have used the integer value  $b = 4$  as the exponential base in the model to describe the NMR spectrum, which leads to the fourth-quarter Cantor set described above. However, it is important to realise that the same mathematical model with a non-integer value of  $b$ , within certain constraints, also leads to a Cantor-type set. In fact, there is no reason why the base  $b$  should be an integer in the model describing the NMR data, and we will show below that the optimal agreement between model and experiment is achieved when  $b = 3.9$ .

The generalization to non-integer bases can also be explained in terms of the expansion given in Eq. S3. Any fractional number can be expressed in that form with a non-integer  $b$  value, using non-negative integer digits  $N_k$  that are less than  $b$  (i.e.  $N_k = 0, 1, \dots, [b]$ ; where  $[b]$  is the integer part of  $b$ ). The theory of number representations in non-integer bases was introduced by the Hungarian mathematician Rényi,<sup>S8</sup> and has recently found interesting applications in the description of quasicrystals.<sup>S8</sup>

One difference with respect to the integer expansion, which in the form given by Eq. S3 can represent any number from 0 to 1, is that the maximum number that the non-integer expansion can represent is  $\sum_{k=0}^{\infty} [b] / b^k = [b] / (b-1)$ . Since for non-integer  $b$  values we have  $b-1 < [b] < b$ , this maximum number is between 1 and 1.5.

We can now define a generalization of the fourth-quarter Cantor set to any base  $3 < b < 4$ , as the set of all  $x$  values which have a  $b$ -expansion (given by Eq. S3), using only the digits  $N_k = 0, 1$  and 2.

Based on the analysis in the previous subsection for the integer-base construction, it is clear that the generalized construction should have the first-level trema between  $(0.3)_b = 3/b$  and  $(0.\bar{3})_b = 3/(b-1)$ , which means that the trema width is  $\frac{3}{b(b-1)}$ .

The three second-level tremas are:

- between  $(0.03)_b = \frac{3}{b^2}$  and  $(0.0\bar{3})_b = \frac{3}{b(b-1)}$ ;
- between  $(0.13)_b = \frac{1}{b} + \frac{3}{b^2}$  and  $(0.1\bar{3})_b = \frac{1}{b} + \frac{3}{b(b-1)}$ ;
- between  $(0.23)_b = \frac{2}{b} + \frac{3}{b^2}$  and  $(0.2\bar{3})_b = \frac{2}{b} + \frac{3}{b(b-1)}$ ;

which are each of width  $\frac{3}{b^2(b-1)}$ . [Note: in order to evaluate periodic expansions in base  $b$  as a function of  $b$ , the following equation is useful:  $\sum_{k=n}^{\infty} (\frac{1}{b^k}) = \frac{1}{b^{n-1}(b-1)}$ ]

By defining tremas in this way *ad infinitum*, only numbers not containing 3 in their base- $b$  expansion will be part of the final set.

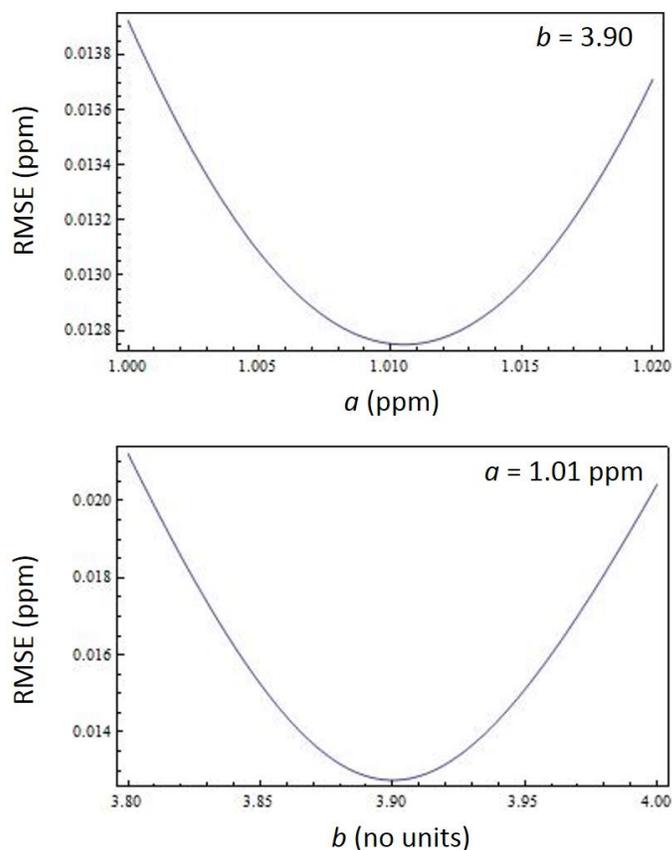
It is clear from these expressions that i) they reduce to those given for the fourth-quarter Cantor set when we set  $b = 4$ ; and ii) that the resulting pattern will also be self-similar with magnification factor  $b$ . It also follows from the above, although it is already obvious from Eq. S3, that the generalized set is not simply “proportional” to the fourth-quarter set. Even if the number of tremas in each step of the construction is the same as in the integer-base case, the relative widths of the tremas with respect to the original interval depends on  $b$ . The exponential base  $b$  is therefore a useful parameter to fit the position of the peaks of our NMR spectrum, which we do in the next subsection.

### 9. 3 Finding the optimal $b$ to fit the NMR data

Finally, we will allow the value of  $b$  in our NMR shielding model to adopt non-integer values and optimize the value to minimize the discrepancy with experimental data. The model (up to septets) is represented by the equation:

$$\Delta\delta = a \sum_{k=1}^3 \frac{N_k}{b^k}$$

where  $a$  is a proportionality constant (which in the main text of the article was shown to be  $\sim 1$  ppm). The fitting of the model involves finding the parameters  $a$  and  $b$  that minimize the root mean square error (RMSE) between model and experimental data.



**Supporting Figure S7.** Variation of the root mean square error (RMSE) between model and experimental data with parameters  $a$  and  $b$ .

Figure S7 illustrates that the minimum discrepancy with experiments is obtained if  $a = 1.01$  ppm and  $b = 3.90$ . This is a shallow minimum, and small variations of  $a$  and  $b$  lead to very small increases in the RMSE. For comparison, the model with approximate values  $a = 1$  ppm and  $b = 4$  as described in the paper, leads to an RMSE that is  $\sim 0.01$  ppm higher than the minimum. This RMSE difference is roughly the same as the precision of the experimental data. Therefore we can report the fitted exponential base parameter as  $b = 3.9 \pm 0.1$ . The difference in RMSE obtained by changing  $a$  from 1 to 1.01 is even smaller. This means that the approximate parameters  $a = 1$  ppm and  $b = 4$  can be considered to be adequate within the precision of the experiment.

## References to Supporting Information

- [S1] Colquhoun, H. M.; Williams, D. J.; Zhu, Z. Macrocyclic aromatic ether-imide-sulfones: versatile supramolecular receptors with extreme thermochemical and oxidative stability. *J. Am. Chem. Soc.* **2002**, *134*, 13346-13347.
- [S2] Parker, M. P.; Murray, C. A.; Hart, L. R.; Greenland, B. W.; Hayes, W.; Cardin, C. J.; Colquhoun, H. M. *Cryst. Growth Des.*, **2018**, *18*, 386–392.
- [S3] Colquhoun, H. M.; Zhu, Z.; Williams, D. J.; Drew, M. G. B.; Cardin, C. J.; Gan, Y.; Crawford, A. G.; Marder, T. B. Induced-fit binding of  $\pi$ -electron-donor substrates to macrocyclic aromatic ether imide sulfones: a versatile approach to molecular assembly. *Chem. Eur. J.* **2010**, *16*, 907-918.
- [S4] Allinger, N. L.; Eliel, E. L. (eds.) *Topics in Stereochemistry*, Vol. 1, John Wiley, New York, **1967**, p.24.
- [S5] Mayo, S. L.; Olafson, B. D.; Goddard III, W. A. DREIDING: a generic force field for molecular simulations. *J. Phys. Chem.* **1990**, *94*, 8897-8909
- [S6] Mandelbrot, B. B.. *The Fractal Geometry of Nature*, **1982**, W. H. Freeman and Co., New York.
- [S7] Edgar, G. *Measure, Topology, and Fractal Geometry*, **1990**, Springer-Verlag, New York.
- [S8] Rényi, A. Representations for real numbers and their ergodic properties, *Acta Math. Acad. Sci. Hung.* **1957**, *8*, 477–493.
- [S9] Burdik, Č.; Frougny, Ch.; Gazeau, J. P.; Krejcar, R. Beta-integers as natural counting systems for quasicrystals, *J. Phys. A* **1998**, *31*, 6449–6472.