

Numerical representations of AB-type copolymer complexes: analysis of ^1H NMR chemical shift patterns in terms of a Smith-Cantor set

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Colquhoun, H. M. and Grau-Crespo, R. ORCID:
<https://orcid.org/0000-0001-8845-1719> (2024) Numerical representations of AB-type copolymer complexes: analysis of ^1H NMR chemical shift patterns in terms of a Smith-Cantor set. *Journal of Mathematical Chemistry*, 62 (7). pp. 1537-1557. ISSN 1572-8897 doi: 10.1007/s10910-024-01614-8 Available at <https://centaur.reading.ac.uk/115839/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1007/s10910-024-01614-8>

Publisher: Springer

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Numerical representations of AB-type copolymer complexes: analysis of ^1H NMR chemical shift patterns in terms of a Smith–Cantor set

Howard M. Colquhoun¹ · Ricardo Grau-Crespo¹

Received: 10 January 2024 / Accepted: 16 March 2024

© The Author(s) 2024

Abstract

When considering the possibility of storing information in the sequence of monomer residues within an AB-type copolymer chain, it is constructive to model that sequence as a string of ones and zeros. The intramolecular environment around any given digit (say a “1”) can then be represented by another string of integers—a *code*—obtained by summing pairs of digits at equivalent positions, in both directions, from that digit. The code can include only integers 0, 1 and 2, and can represent a number in any base b higher than 2. In base $b=3$ the resulting set of codes includes *all* numbers (because only digits 0, 1 and 2 occur in ternary expansions), but in any base $b>3$ the codes define a limited set of numbers comprising a fractal we term a Smith–Cantor set. The ^1H NMR spectrum of a random, AB-type co(polyester-imide) shows, on complexation with pyrene, a pattern of complexation shifts approximating very closely to the Smith–Cantor set for which $b=4$. Other co(polyimide) complexes show a ^1H NMR pattern corresponding to a specific *sub-set* of this fractal. The sub-set arises from a “stop-at-zero” limitation, whereby digits in the initial string are set to zero for code-generating purposes if they occur beyond a zero, as viewed from the central “1”. The limitation arises in copolymers where pyrene binds by intercalation between *pairs* of adjacent diimide residues. This numerical approach provides a complete, unifying theory to account for the emergence of fractal character in the ^1H NMR spectra of AB-type copolymer complexes.

Keywords Fractals · NMR spectroscopy · Ring-current shielding · Linear copolymers

✉ Howard M. Colquhoun
h.m.colquhoun@reading.ac.uk

¹ Department of Chemistry, University of Reading, Whiteknights, Reading RG6 6DX, UK

1 Introduction

The processing of information in biological systems is achieved, universally and with very high precision, through the enzyme-catalysed chemistry of a set of sequence-defined, high molecular-weight, linear copolymers (DNA and/or RNA, and proteins) [1–3]. In principle, however, *any* copolymer sequence can represent information, because even the simplest AB copolymer is the logical equivalent of a binary string [4, 5]. Some significant progress in developing a synthetic “information chemistry” has been made in recent years, notably with the discovery of sequence-specific polymerisation methodologies and mass-spectrometric sequencing techniques [6–8], information-transfer protocols [9–11], and the use of small “reader-molecules” to recognise copolymer sequence-information [12–18].

We have recently shown that highly sequence-dependent ^1H NMR complexation-shifts are produced in the spectra of certain random AB-type copolyimides based on 1,4,5,8-naphthalene tetracarboxylic diimide (NDI) on complexation by an aromatic “probe” molecule such as pyrene or perylene [15–17]. This phenomenon arises from cumulative ring-current shielding [19, 20] of the central residue in an NDI-centred sequence, produced by the probe-molecules complexing to NDI residues by electronically complementary π – π -stacking [21, 22]. Such shielding results not only from the probe-molecule binding directly at the central “observed” NDI residue, but also from complexation to NDI residues at *neighbouring* (and next-neighbouring, and next-next-neighbouring etc.) positions viewed in both directions from the centre of the sequence. Copolymer-to-probe binding evidently operates under fast-exchange conditions on the NMR timescale, since separate resonances corresponding to bound and unbound NDI residues are not observed.

Multiple NDI signals from the copolymer are seen at high pyrene concentrations, even under fast-exchange conditions, because each observed NDI residue is at the centre of a separate copolymer sequence. This “surrounding sequence”, which defines the intramolecular environment of the central observed NDI residue may be of any length, although spectroscopic resolution generally limits the maximum observable length to a heptad or nonad. Note that NDI residues at the centres of different sequences are inherently distinct and thus give different complexation shifts in the presence of pyrene, regardless of fast exchange between pyrene-bound and unbound states.

The concept of a central “observed” residue in a given sequence is important, because it greatly simplifies sequence-analysis in high molecular weight copolymers. Even though a given copolymer sequence may contain many observable monomer residues (in the present context, NDI), each of these is also at the centre of its *own* sequence, overlapping with the original sequence while still, in NMR terms, representing a specific, intramolecular environment. Consequently, the central residue can be treated separately from other chemically equivalent residues in the same sequence. In NDI-based copolymers, ring-current shielding resulting from complexation of an aromatic molecule such as pyrene *amplifies* the differences between

proton magnetic environments in a copolyimide chain and enables the assignment of specific NMR resonances to different comonomer sequences [15].

Our work on the supramolecular chemistry of binary co-polyimides containing the 2,4,5,8-naphthalenediimide (NDI) unit has also shown that complexation with probe-molecules such as pyrene results in their ^1H NMR spectra displaying self-similar patterns related to certain classes of fractal [15–17]. These patterns were accounted for in terms of a physical model whereby the degree of shielding of a central, observed NDI residue by pyrene falls away exponentially as neighbouring NDI binding sites are located further and further out, in both directions, from the centre.

In a copolyimide containing two different types of diimide residue, one (NDI) strongly binding for pyrene and the other strictly non-binding, we can assign the digit 1 to NDI and 0 to the other. To define the intramolecular environment of any given NDI residue—a key factor in determining the chemical shift of its protons in ^1H NMR—we can then *sum* digits at equivalent positions in the resulting string, to give a new set of digits whose values are allowed to fall off exponentially on reading from left to right. This new set of digits is then the *exact equivalent of a number*, with the fall-off factor corresponding to the base in which the number is defined (e.g. 10 for decimal, 8 for octal, 4 for quaternary etc.). On this basis, we have developed and here report a new and purely numerical approach to understanding the observed, fractal NMR patterns. This novel approach involves modelling the intramolecular environment of any monomer residue in a two-monomer, high molecular weight copolyimide as a digital sequence within a single, infinite, two-integer string.

2 Results and discussion

2.1 The numerical environment of an integer in a two-component numerical string

Consider an infinite, random sequence of ones and zeros, in which the probability that any given digit is a one (or a zero) is one half. Such a sequence must contain equal numbers of the integers 1 and 0, and a short segment of such a random sequence might be:

$$\infty \dots 01010001101110101011100010110 \dots \infty \Rightarrow (14 \text{ zeros and } 15 \text{ ones})$$

Any integer (say a **1**) in such a sequence is, by definition, located at the junction of two sequences that extend outwards from itself, backwards and forwards, to infinity. Here we use a bold character to distinguish the central **1** from all the other 1 s in a given sequence. For example, in the above:

$$\infty \dots 01010001101110 \leftarrow \mathbf{1} \rightarrow 01011100010110 \dots \infty$$

To describe, numerically, the *environment* of the chosen integer **1**, we identify pairs of digits at equivalent positions by reading outwards in both directions from this integer. Summing each such pair of digits (*not* identified as binary digits, for reasons outlined below) can give only one of three possible results, i.e. $[0+0]=0$,

$[0+1]$ or $[1+0]=1$, and $[1+1]=2$. Summing integers at equivalent positions then affords a new sequence which we will refer to as the *code* for the central integer **1**. Clearly, every integer 1 in the original sequence can be assigned such a code, which is then a numerical descriptor for the environment of the integer in question: the integer N_k at position k in the code gives the number of 1 s at the numerical “distance” k . Thus, for the integer **1**, centred in the sequence above, the resulting code is:

$$02121210011120\dots\infty$$

Clearly, no integer higher than 2 can occur in this type of code, because there can only be a maximum of two 1’s at a given numerical distance from the position of interest. Note that there are two ways in which the digit 1 can arise in a code ($0+1$ or $1+0$), so different original sequences can give rise to the same code, which would then have some degree of degeneracy (see Sect. 2.3). Like the original sequence, each code extends to infinity. But, unlike the original string, the code has a definite starting point, i.e. the sum of the digits at the pair of positions immediately adjacent to the “central” **1**. The code also has a definite “reading” direction, i.e. left to right, corresponding to increasing distance of the original digits from the central **1**.

The code is obviously not a binary number, since the digit 2 does not appear in binary numbers, but it could be a number in any higher base b (ternary, quaternary etc.). If the code is to be seen as a number in a base b , then successive digits N_k must take values reduced by a factor b^k as the code is read from left to right. The exponent k is determined by the position of each digit in the code, the first digit being assigned the exponent $k=1$, the second $k=2$ and so on. The values of the code-digits are then summed to give a total (decimal) value, T , for the code:

$$T = \sum_{k=1}^{\infty} \frac{N_k}{b^k}. \quad (1)$$

Here we are choosing to read the base- b numbers as fractions of one, for reasons that will become clear below (although, in principle, numbers greater than one can also be represented by Eq. 1, if negative values of k are allowed). Since the term b^k occurs in the denominator of each term of this summation, the values of successive digits fall off exponentially by the factor b . In the codes defined above, the only digits assignable to N_k , for any value of k , are 0, 1 or 2, so the minimum value of T in any base is zero (when N_k is always zero) and the maximum value (when N_k is always 2) is given by:

$$T_{\max} = \sum_{k=1}^{\infty} \frac{2}{b^k} = \frac{2}{b-1} \quad (2)$$

Thus, the range of code values assigned as ternary ($b=3$) numbers is 0 to 1; for quaternary numbers ($b=4$) it is 0 to $2/3$ (or 0.666, using a notation that underlines the repeating terminal digit); for quinary numbers ($b=5$) it is 0 to 0.5; and so on.

2.2 The code as a Smith–Cantor set

For $b=3$ and $k_{\max}=\infty$, *all* the numbers in the range 0 to 1 are described by Eq. 1 (i.e. the set is everywhere dense) because the only values of N_k required in the ternary system are 0, 1 and 2. However, for bases higher than 3, Eq. 1 describes just a restricted set of numbers, since numbers expressed in these higher bases may require integers N_k higher than 2. For example, a quaternary representation ($b=4$) of all fractional numbers between 0 and 1 requires the digits 0, 1, 2 and 3. If the digit 3 is not allowed, then Eq. 1 describes only a reduced set of all fractions. Sets of this type, where at least one digit required for the b -based representation is forbidden, are nowhere dense.

The most famous example of such a set is the fractal known as Cantor’s “middle-third” set [23, 24], which can be defined as the set of numbers between 0 and 1 which have a ternary ($b=3$) representation *not including the integer 1*. In the codes we have defined above, for bases b higher than 3, there are always “missing” integers in the set of numbers generated by Eq. 1 (because integers higher than 2 are always forbidden). Equation 1, with $b>3$, then also defines a fractal: a “last-fraction” set. Such a set was first described by Smith in 1875 [25], in paper that probably gave the first mathematical description of *any* fractal and preceded Cantor’s work [26] by almost a decade, even if Cantor’s name is now more generally attached to such sets [27–29]. Because of this historical anomaly, fractals defined as *last-fraction* sets will hereafter be referred to as Smith–Cantor sets.

Cantor’s “middle-third” set [26] can be constructed graphically (Fig. 1a) by taking a line of length one, dividing it into three and deleting the middle third (equivalent to removing ternary numbers containing the integer 1 in the first position); then deleting the middle third from the two remaining segments (equivalent to removing ternary numbers containing 1 in the second position) and repeating these operations indefinitely [30]. In the limit, this construction is equivalent to the above definition of the set as “all the numbers in the range 0 to 1 which admit a ternary representation not including the integer 1” [30].

Analogously, the graphical construction of a Smith–Cantor *last-fraction* set involves taking a line of length one, dividing it into b segments, deleting the *last* (i.e. right-hand) segment or segments, and iterating this procedure indefinitely. When carried out for $b=4$, thus removing quaternary numbers containing the digit 3, we generate the “fourth-quarter” Smith–Cantor set. The first four iterations of its construction are shown in Fig. 1b.

In the present context, the codes describing the numerical environment of every integer 1, in an infinite sequence of ones and zeros, can thus represent either the set of *all* ternary numbers ($b=3$) or, for $b>3$, a *restricted set* of numbers comprising a last-fraction (Smith–Cantor) set. Thus, setting $b=4$ and $N_k=0, 1$ or 2 (i.e. excluding 3), generates the fourth-quarter Smith–Cantor set. For higher values of b , and with N restricted to these same three values, Eq. 1 generates sets with greater numbers of disallowed integers. For example, with $b=5$, quinary numbers containing the integers 3 and 4 are forbidden, so that in geometrical terms the result might be termed the “last two-fifths” set. Similarly, for $b=6$ we would arrive at the “last three-sixths” set, whose construction is shown in Fig. 1c.

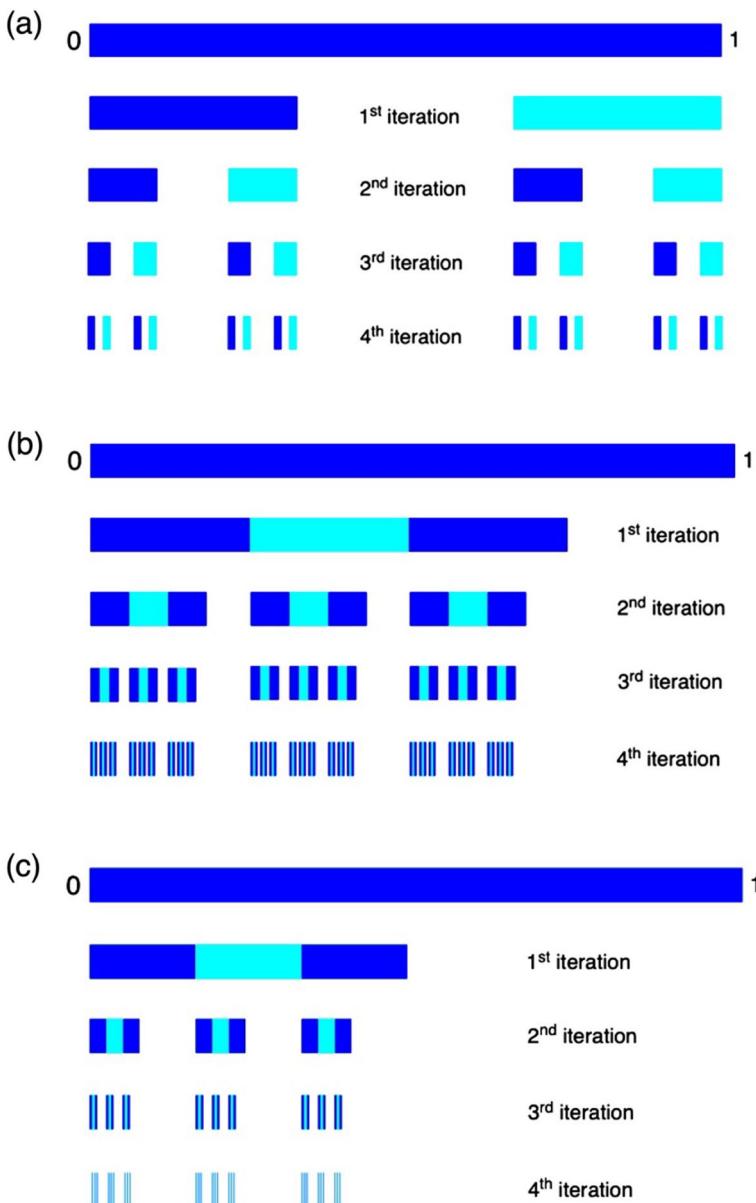


Fig. 1 Graphical constructions (first four iterations) of **a** the “middle-third” Cantor set; **b** the “fourth-quarter” Smith–Cantor set, and **c** the “last three-sixths” Smith–Cantor set. Colours are used to differentiate the segments defined at each step of the construction (Color figure online)

Comparison of the fourth-quarter and last-three-sixths constructions suggests that last-fraction Smith–Cantor sets for which N_k is limited to 0, 1 and 2 have a qualitatively similar structure, with numbers grouped in triplets of triplets of triplets etc.,

but also that the *numerical distribution patterns* of the numbers are very different for different values of b . Thus, were a “last fraction” pattern of this type to be encountered in the physical world, the identity of b , and thus of the associated Smith–Cantor set, could be determined by measuring the relative separations of the observed data points [17] (see discussion of experimental data in Sect. 2.3).

As noted earlier, a code defining the environment of an integer in an infinite two-digit sequence is itself infinite. However, because successive terms of the summation shown in Eq. 1 decay exponentially in value, the converged sum may be approximated by truncation after only a small number of terms. For example, assuming that the infinite code:

$$0212121001\dots\infty$$

represents a quaternary fraction, its value, from Eq. 1, would be:

$$T = 0/4^1 + 2/4^2 + 1/4^3 + 2/4^4 + 1/4^5 + 2/4^6 + 1/4^7 + 0/4^8 + 0/4^9 + 1/4^{10} + \dots \infty$$

Summing these first ten terms gives (to eight decimal places) $T=0.14996435$. However, if the sum is truncated after the sixth digit, the value of T changes to 0.14990237, i.e. with a difference only in the fifth decimal place. Even truncating after the third digit gives a still-useful approximation for T (0.1484375). Note that a three-digit code (here 021) arises from pairwise summation of the “outer” six digits of a seven-digit sequence, since the central digit does not feature in the numerical description of its own environment. Also, truncating all **1**-centred sequences in an infinite sequence to **1**-centred heptads reduces the number of different possible environments for the integer **1** from infinity to just 2^6 , i.e. 64, as there are only two allowed digits (0 or 1) at each of six positions around the central **1**. Moreover, these 64 sequences give rise to only 27 different codes, since the summation process that converts heptad sequences to codes reduces the number of “environment” digits from six to three, at the same time increasing the number of possible integers to three (0, 1 or 2) at each of the three positions, meaning there are just 3^3 (i.e. 27) possible code-combinations.

It will be noted that, in the third iteration, the graphical construction of the fourth-quarter Smith–Cantor set shown in Fig. 1b also (and equivalently) generates a total of 27 segments. Indeed each successive iteration of this construction is strictly equivalent to extending the numerical description of the environments of the all the 1s in a sequence by one digit in each direction. Thus the first iteration corresponds to a one-digit code, representing the sum of the two digits immediately adjacent to a “central” **1**, in a three-digit sequence. The second iteration corresponds to a two-digit code and thus to a five-digit, **1**-centred sequence and, as we have seen, the third iteration corresponds to a three-digit code for a seven-digit sequence of this type.

All the sets defined by Eq. 1 (with $N_k \leq 2$) are self-similar, in the sense that if we zoom in on a part of the set, the magnified portion resembles the whole set. They are also fractals, in the definition coined by Mandelbrot [31], with a non-integer fractal dimension D which can be calculated as:

$$D = \frac{\ln(\text{number of allowed integers in } \{N_k\})}{\ln(b)} \quad (3)$$

The fractal dimension of the traditional middle-third Cantor set is therefore $D = \ln 2 / \ln 3 \approx 0.631$, whereas for the fourth-quarter set $D = \ln 3 / \ln 4 \approx 0.792$. In the limiting (non-fractal) case when all integer digits from zero to $b - 1$ are allowed in the expansion, we recover the topological dimension of a line, $D = 1$.

It should also be noted that, in the definition of these fractal sets, there is no requirement for b to be an integer (except that sets defined for integer b values are easier to represent using graphical constructions, as shown in Fig. 1). For non-integer values of b , Eq. 1 can still be used to express any fractional number 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 is well developed [32] and has found interesting applications in the description of quasicrystals [33].

2.3 Code values as signatures of local environment: degeneracy and frequency of occurrence

In order to attach generalised physical meaning to code values, we can now say that a code value is a local signature for any site, that results from contributions from all “occupied” neighbours (1s in the original string) of that site, on condition that such contributions decay exponentially with the numerical distance k from the neighbour to the site. Equation 1 can be written more generally as:

$$T = \sum_{k=1}^{\infty} N_k \exp(-\beta k) \quad (4)$$

where $\beta = \ln b$. As long as the decay is fast enough ($\beta > \ln 3$) the collection of all the signatures, corresponding to all the possible local environments in the original string, forms a Smith–Cantor set.

We now give consideration to the frequency with which a given code value or signature appears. Because the digit 1 can appear in a code as a result of two different pairs $[0 + 1]$ or $[1 + 0]$, each occurrence of the digit 1 in the code introduces a degeneracy of 2, whereas the 0’s or 2’s in the code do not introduce degeneracy. Therefore, the total degeneracy, Ω , can be expressed as a function of the code digits N_k :

$$\Omega = \prod_k 2^{\delta_{1,N_k}} \quad (5)$$

where the symbol \prod_k denotes a product of the argument over all the values of k , and $\delta_{i,j}$ is the Kronecker delta, defined as:

$$\delta_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

The expression above simply means that the degeneracy doubles for each integer “1” in the code, because there are two ways of achieving an occupancy of 1 at each position.

If the initial string of ones and zeros is fully random, then the frequency with which a given code, (characterising a specific local environment) appears in the set of all codes, is simply proportional to the degeneracy of the code. The maximum frequency corresponds to codes containing only ones (e.g. 111 for 3-digit codes). Relative to that maximum, the frequency of any code corresponds to:

$$\frac{\Omega}{\Omega_{\max}} = \frac{1}{\prod_k 2^{1-\delta_{1,N_k}}} \quad (6)$$

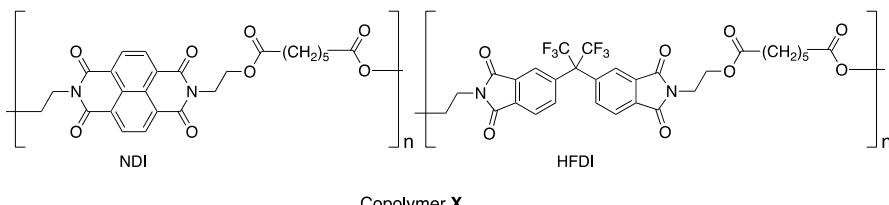
so the frequency of appearance of a code is halved for each code-digit different from 1. Thus, the peak with code 101 has a relative intensity of $\frac{1}{2}$ because it has one digit different from 1, whereas the peak with code 100 has a relative intensity of $\frac{1}{4}$ because it has two digits different from 1.

The discussion above implies that if it were possible to measure signatures of local environments in any physical realisation of a linear two-component string, the result (as long as the exponential decay of contributions to the signature was fast enough) would be a fractal set with intensities modulated by Eq. 5. In what follows we describe some physical realisations of this idea in the context of ^1H NMR spectra of complexes of random, two-component copolyimides. These examples, which were first presented in Refs. [15–17], are discussed here within the more general numerical framework introduced above, thus increasing our understanding of such systems and perhaps facilitating the discovery of other physical realisations of the numerical model.

2.4 Experimental results

2.4.1 ^1H NMR spectra of NDI-based copoly(ester-imide)s

Any AB-type copolymer of sufficiently high molecular weight can be approximated numerically as an infinite string, in which the two different co-monomers are assigned as digits one and zero. As noted in Sect. 2.1, there is no requirement for these to be *binary* digits—they can represent digits in ternary, quaternary, or indeed any other number system. An AB copolymer central to the present discussion (copolymer **X**) is shown below. It contains equimolar amounts of the two different diimide units, and their distribution within the chain is essentially random [17].



Copolymer **X**

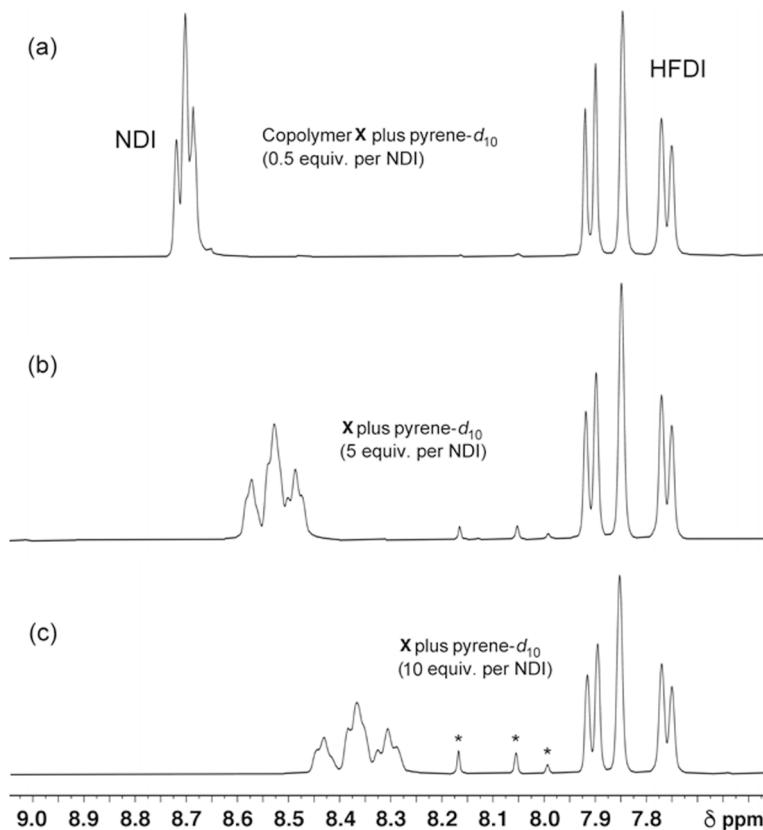


Fig. 2 The ^1H NMR spectra of copolymer **X** (4 mM in NDI residues in CDCl_3 /trifluoroethanol, 6:1 v/v) in the presence of increasing levels of pyrene- d_{10} . At low levels of pyrene (0 to 3 equivalents per NDI) only three NDI resonances are seen [spectrum (a)], assignable to the three possible NDI-centred triad sequences. At higher pyrene concentrations nine resonances are resolved [spectra (b) and (c)]. Starred resonances arise from residual protons present in the (99.8%) deuterated pyrene [17]

In terms of a numerical description, the 1,4,5,8-naphalenediimide (NDI) unit might be identified with the digit 1 and the hexafluoroiso-propylidenediimide (HFDI) unit with the digit 0. Successive diimide units are invariably linked by identical aliphatic-diester units, so these can be ignored in any digital representation of the copolymer. Experimentally, the ^1H NMR spectrum of this copolymer in the diimide region (Fig. 2) shows two groups of resonances, a narrow 1:2:1 pattern at around 8.7 ppm, assigned to the NDI protons, and a more complex pattern at higher field assignable to the HFDI protons. Focusing just on the three NDI resonances, these may be assigned to the triad sequences [HFDI-NDI-HFDI, [NDI-NDI-HFDI or HFDI-NDI-NDI], and [NDI-NDI-NDI], or in digital notation [010], [110 or 011], and [111]. In each case we can assign the observed resonance to just the *central* NDI unit of the triad because any outer NDI residues, being part of a longer copolymer chain, are themselves at the centres of *other*

triad sequences and can thus be treated separately. Moreover, in the two unsymmetrical triad sequences ([NDI-NDI-HFDI] and [HFDI-NDI-NDI]) the intramolecular environments of the central NDI units are the same because these two triads are simply mirror images of one another. In NMR terms, such sequences are *degenerate*, i.e. they give resonances with identical chemical shifts: the resulting NMR signal thus has twice the intensity of resonances from the other two (non-degenerate) triad sequences.

Aromatic π -donor molecules such as pyrene and perylene are well known to form noncovalent complexes with NDI residues, via π – π donor–acceptor interactions, resulting in ring-current shielding of the NDI protons and a consequent upfield shift of the corresponding resonances [34, 35]. In the case of copolymer **X**, progressive addition of pyrene (perdeuterated to avoid resonance-overlap) to a solution of this copolymer results in upfield shifts of all three signals corresponding to the NDI-centred triad sequences discussed above. In addition, as shown in Fig. 2, the three NDI resonances (initially between 8.60 and 8.80 ppm) not only *shift* but also *resolve*, ultimately into nine resonances when the molar ratio of pyrene to NDI units in the copolymer reaches 10:1. Conversely, resonances associated with the non-binding HFDI residues, in the range 7.70 to 8.00 ppm, are entirely unaffected [17].

As noted above, pyrene binding to NDI in this system occurs under fast-exchange conditions on the NMR timescale, because separate resonances corresponding to bound and unbound NDI protons are not observed. Note that the resolution of the three initial NDI resonances into three “triplets” closely parallels the progression from iteration 1 to iteration 2 of a last-fraction Cantor set, as defined by Eq. 1 (Fig. 1b, c). On this basis, analysis of the relative separations, in ppm, between the nine resonances, as the pattern evolves, allows an *experimental* value for b in Eq. 1 to be determined. This value emerges [17] as being very close to the integer 4, suggesting that in this system the corresponding sequence-codes are quaternary numbers and thus that the pattern of chemical shifts for the NDI protons in the pyrene complex of copolymer **X** may be identified with the fourth-quarter Smith–Cantor set (Fig. 1b).

Given the analogy between copolymer **X** and a random, infinite, sequence of ones and zeros, our next task was to simulate the experimental chemical shifts in terms of the fourth-quarter Smith–Cantor set. This required the inclusion of an additional variable in Eq. 1, to allow for the variation of total shielding T with the concentrations of copolymer and pyrene. Thus, the sum of shieldings T may be scaled linearly by a factor a that depends on the molar ratio of pyrene to NDI and on the concentration of NDI residues. This scaling factor reflects an increasing level of ring-current shielding with (i) an increasing overall concentration of the copolymer/pyrene system, where a higher concentration tending to shift the binding equilibrium towards the bound state, and/or (ii) an increasing molar ratio of pyrene to NDI residues, with a higher ratio leading to a higher proportion of NDI residues being in the bound state). The factor a is assigned the units of ppm and this enables the otherwise

dimensionless total-shielding factor, T , to be expressed as a predicted complexation shift for the central, “observed” NDI residue in each sequence:

$$T = a \sum_{k=1}^{k_{\max}} \frac{N_k}{b^k}. \quad (6)$$

Equation 6, with $a=1$, $b=4$, $N_k=0$, 1 or 2, and $k_{\max}=\infty$, is the mathematical definition of the fourth-quarter Smith–Cantor set [15, 17, 25, 28]. Although a obviously changes *between* spectra as the pyrene concentration changes (Fig. 2), it is a *constant* for each individual spectrum, and each spectrum can therefore be predicted from Eq. 1. Since fractals are scale-invariant, the introduction of the factor a does not affect the fractal nature of the system.

Finally, Eq. 1 can be expanded, taking the ring-current shielding by pyrene bound directly to the central NDI (T_0) out of the summation, as this shielding is always present whatever the sequence under consideration (Eq. 7).

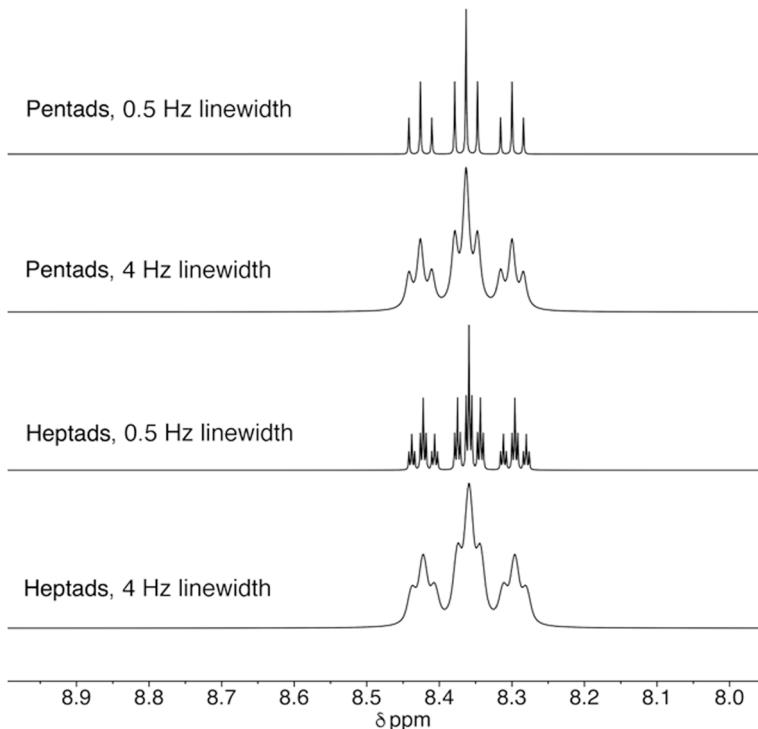


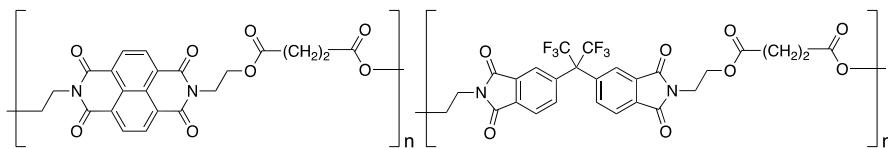
Fig. 3 Simulated ^1H data, from Eqs. 5 and 7, for copolymer **X** in the presence of pyrene (10 equivalents), based on pentads (above) and heptads (below) [17]. The simulations should be compared with the experimental spectrum shown in Fig. 2c. Simulations were carried out using the “Peak Table to Spectrum” script within the NMR software package *Mestrenova*, version 14.1.1 (Mestrelab Resarch, Santiago, Spain). This tool employs a generalised Lorentzian lineshape, and inputs are the values for complexation shifts (ppm) and intensities (degeneracies) predicted by the present theory, together with a user-defined linewidth

$$T = T_0 + a \sum_{k=1}^{k_{\max}} \frac{N_k}{b^k}. \quad (7)$$

Since the integrated intensities of ^1H NMR resonances normally correlate with the relative proportions of the different types of proton in the molecule being studied, it is predicted that the relative intensities of the different NDI resonances, representing different NDI-centred sequences in the copolymer, will be equal to the relative degeneracies of these sequences as given by Eq. 5.

Thus we now have a complete numerical model, represented by Eqs. 5 and 7, for predicting the development of the ^1H NMR spectrum (in the NDI region) of a random, equimolar, two-component copolyimide (e.g. copolymer **X**), in the presence of an increasing concentration of an aromatic-probe molecule such as pyrene. A predicted set of spectra (for $k_{\max} = 2$, i.e. considering pentad sequences) using a physically-reasonable linewidth of 4 Hz is shown in Ref. [17] where it is compared with a corresponding set of experimental spectra for pyrene:NDI molar ratios in the range 3 to 10. Although the comparison is good, simulation using a longer sequence-length (heptads rather than pentads) reproduces the experimental signals even more closely as a result of the emergence of additional fine structure (Fig. 3) at the next iteration of the Smith–Cantor set. This produces twenty-seven lines instead of nine, but at the realistic linewidth of 4 Hz only nine lines are evident in the simulation and the resulting resonance-profile is a much better match to experiment (Fig. 2c) [17].

The ^1H NMR spectra of copolymers analogous to copolymer **X**, but in which the length of the polymethylene spacer unit $[\text{CH}_2]_n$ was varied from $n=1$ to 8, were also investigated in terms of their binding to pyrene [16]. For values of n in the range 6 to 8, the results were very similar to those described above for $n=5$, but for shorter spacer-units, notably the copolymer with $n=2$ (copolymer **Y**), completely different NMR behaviour was observed. Specifically, for any given ratio of pyrene to NDI there is a ca. threefold increase in the complexation shifts for the NDI protons, signifying much stronger binding to pyrene, and a very different resonance pattern emerges at the higher pyrene concentrations [16].

Copolymer **Y**

Atomistic simulations indicate that these differences result from a change in pyrene-binding behaviour, with the aromatic molecule now being bound very strongly by *intercalation* between neighbouring NDI units. The spacer-unit with $n=2$ was found to provide a highly favourable chain-fold geometry for such “dual-site” binding, and a new binding-model was thus developed in which the much weaker single-site binding seen for values of n in the range 5 to 8 could be ignored [16].

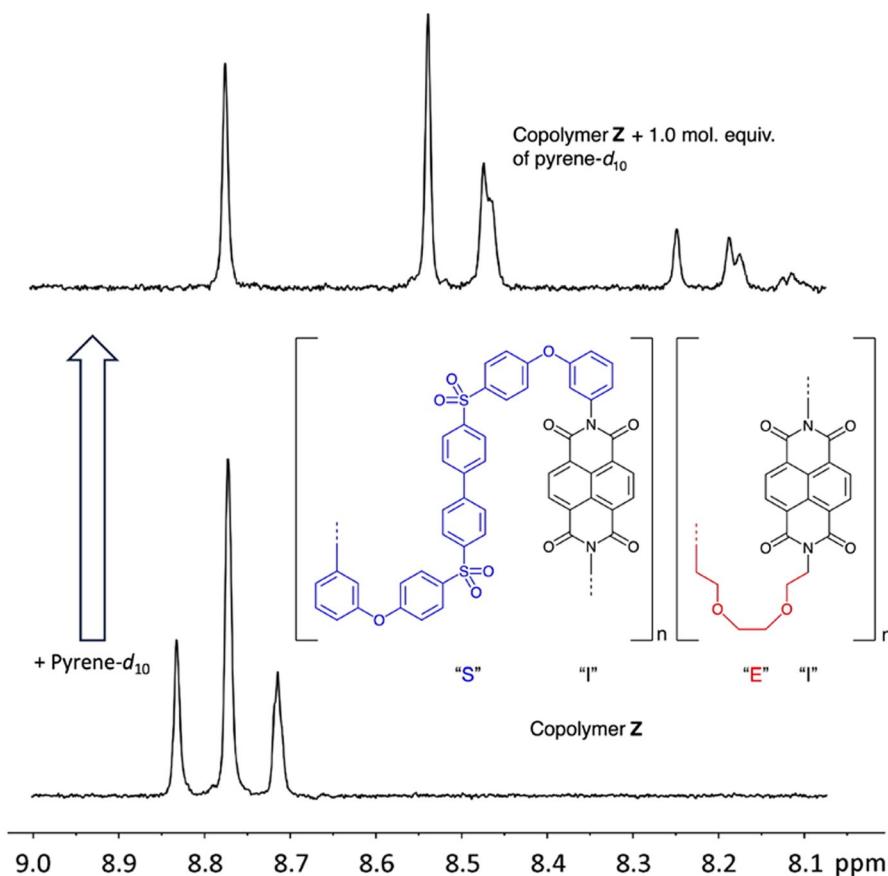


Fig. 4 The ^1H NMR spectra (diimide region) of copolymer **Z**. The lower spectrum is of the pure copolymer. The upper spectrum is of the same copolymer in the presence of 1 mol-equivalent (per NDI) of pyrene- d_{10} [15] (Color figure online)

2.4.2 ^1H NMR spectra of an NDI-based copoly(ether-sulfone-imide)

Investigation of a further range of potentially chain-folding copolyimides led to the discovery of a copoly(ether-sulfone-imide) (**Z**) for which even stronger pyrene-binding was observed [15]. On complexation with pyrene, copolymer **Z** (Fig. 4) gave a pattern of NDI resonances similar to—but much more highly-resolved than—that observed for copolymer **Y**. Once again, the spectra are consistent with fast-exchange of pyrene between the bound and unbound states.

For copolymer **Z**, computational modelling showed that the triethylene-dioxy linker “E” (see Fig. 4) brings two adjacent NDI residues into an *extremely* favourable geometry for intercalation of pyrene between two NDI units. This was confirmed experimentally by single-crystal X-ray analysis of an analogous oligomer-complex of pyrene [15]. Such intercalative or dual-site binding of pyrene to copolymer **Z** again leads to a quite different NDI-resonance pattern from the fourth-quarter

Smith–Cantor set observed for copolymers showing single-site binding (Fig. 3). The dual-site binding pattern (Fig. 4) is no longer obviously fractal, but it retains an element of self-similarity [15]. Graphical analysis shows that this spectrum is made up of several smaller copies of itself, scaled at 1/4, 1/16 and 1/64th, translated to upper limit of the original pattern and then recombined [15]. Self-similarity is now present about only a single point—the upper limit of the spectrum—rather than about an infinity of points as in a complete Smith–Cantor set. The dual-site binding pattern is thus analogous to a logarithmic spiral, which is self-similar only about its origin.

Nevertheless, the observed scaling-factor of 1/4 strongly suggests that the “dual-site” resonance pattern is again somehow related to the fourth-quarter Smith–Cantor set and indeed, as shown in Sect. 2.5, it is found to be a sub-set of the latter, defined by the introduction of just a single limitation in the assignment of sequence-codes.

2.5 A Smith–Cantor sub-set: the “stop-at-zero” limitation

Copolymer **Z** (Fig. 4) can be represented digitally as a random string of one and zeros by disregarding the NDI units—one of which is present between every two adjacent diamine residue and so contributes no sequence-information. We then assign the triethylene-dioxy residue “E” as digit 1 and the diether-disulfone residue “S” as digit 0. A segment of this copolymer might then be represented as either –ESSEEESSESSESESE– or –100100011011010010–. This formulation of the copolymer (a much simpler and more productive representation than that described in Ref. [15]) is especially valuable because every “E” links two NDI units, giving a tightly chain-folded binding site for intercalation of pyrene. Thus “E” (or 1) represents a strongly pyrene-binding position in the copolymer chain. Conversely, every “S” *also* links two NDI units, but now with widely-spaced diimide units, so that “S” (or 0) can be regarded as effectively non-binding in this situation.

Using the digital representation of copolymer **Z** described above, we have discovered that a pattern of codes consistent with the spectrum shown in Fig. 4 can be generated merely by halting the sequence-reading process (in either direction from

Table 1 Implementation of the “stop-at-zero” limitation for codes based on 1-centred heptad sequences

Heptad sequence	Code	Decimal value	Degeneracy
ΘΘ010ΘΘ	000	0	$2^4=16$
ΘΘ0110Θ and Θ0110ΘΘ	100	0.25	$2^3=8 \times 2=16$
ΘΘ01110 and 01110ΘΘ	110	0.3125	$2^2=4 \times 2=8$
ΘΘ01111 and 11110ΘΘ	111	0.328125	$2^2=4 \times 2=8$
Θ01110Θ	200	0.5	$2^2=4$
Θ011110 and 011110Θ	210	0.5625	$2^1=2 \times 2=4$
Θ011111 and 111110Θ	211	0.578125	$2^1=2 \times 2=4$
0111110	220	0.625	$2^0=1$
0111111 and 1111110	221	0.640625	$2^0=1 \times 2=2$
1111111	222	0.65625	$2^0=1$

any integer 1) once a zero is reached. We will refer to this as the “stop-at-zero” limitation. Then, as before, integers at equivalent positions relative to the “central” 1 are summed in pairs to give the code for the corresponding sequence.

For example, considering 1-centred, two-digit, *heptad* sequences there are, as shown earlier, 64 (i.e. 2^6) of these, giving rise to just 27 (3^3) three-digit codes. However, if the “reading” process is terminated once a zero is reached, then all the integers *beyond* the zero are treated, for code-generating purposes, as zeros. As shown in Table 1, applying this “stop-at-zero” limitation to heptad sequences results in the generation, from Eq. 1, of only ten (rather than twenty-seven) different codes, consistent with the ten resonances seen in Fig. 4. The symbol Θ used in Table 1, occurring beyond the point where a zero is reached, indicates that the integer at that point in the sequence can be either 1 or 0. In either case, symbol Θ then counts for code-generating purposes as a zero.

As noted above, the observed scaling-factor of 1/4 [15] strongly suggests that the codes in Table 1 are again quaternary numbers, and they are hereafter treated as such. Their decimal equivalents are obtained from Eq. 1 by setting $b=4$, $N_k=0, 1$, or 2, and $k=1$ to 3. The degeneracy of each code is equal to the number of different heptad sequences that give rise to that code. Because of the “stop-at-zero” limitation outlined above, degeneracies are no longer given by Eq. 4. However, a simple calculation of each degeneracy is shown in Table 1, where there are two possibilities (1 or 0) for each position Θ . For example, a sequence containing three positions of type Θ has a degeneracy of 2^3 (=8), but if the sequence is unsymmetrical then it also has an

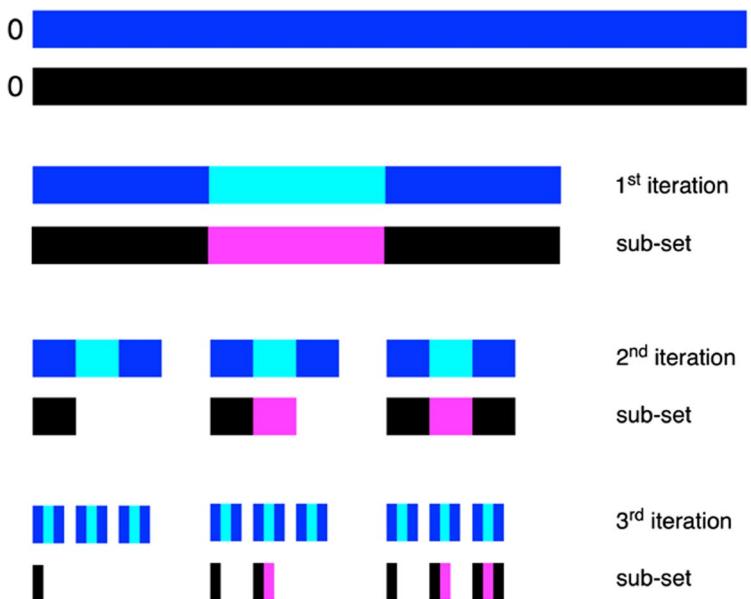


Fig. 5 Graphical construction of the Smith–Cantor fourth-quarter set (light and dark blue) and of the sub-set arising from the “stop-at-zero” limitation (magenta and black) (Color figure online)

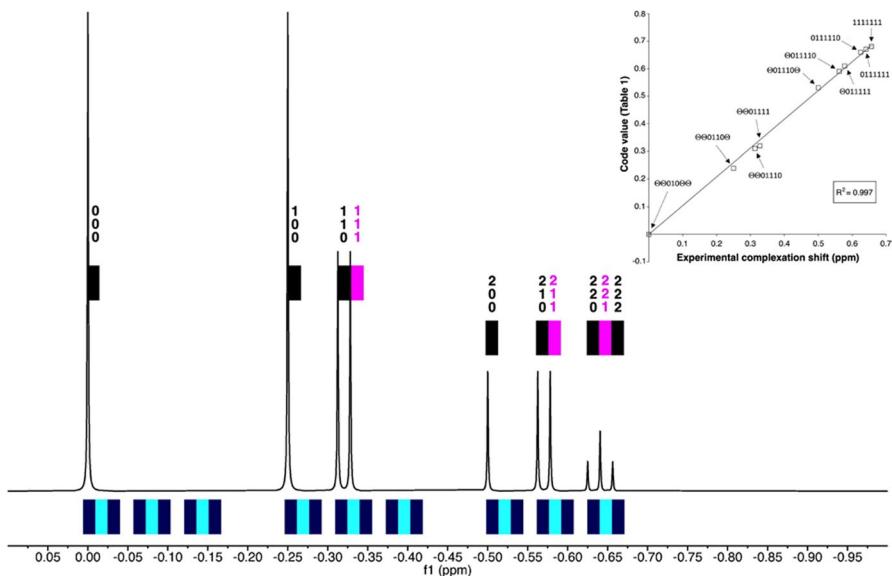


Fig. 6 Simulation of the ^1H NMR spectrum of the 1:1 complex between copolymer **Z** and pyrene, using only the numerical data (range 0 to 1) given in Table 1, at a simulated linewidth of 0.5 Hz. Comparison with the experimental spectrum shown in Fig. 4 indicates remarkably good agreement in relative peak positions and intensities. Indeed, a correlation plot (inset) between experimental complexation shifts and the code values shown in Table 1 is essentially linear, with an agreement factor, R^2 , of 99.7% (Color figure online)

“environmentally-equivalent” (in an NMR context) reverse-sequence, giving a total degeneracy of $2^3 \times 2 = 16$ (cf. line 2 of Table 1).

As seen in the experimental spectrum (Fig. 4), the single self-similarity point of the sub-set is located at its upper limit as given by Eq. 2, i.e. 0.666 for quaternary numbers. The numbers in the sub-set are obviously defined by the same equation as the full set (Eq. 1), but only a selected group (Table 1) are permitted by the “stop-at-zero” limitation. The resulting sub-set of codes is obtained very simply (as above) by applying this limitation to the code-generating process, but it is also possible to find rules for an iterative graphical construction of the fourth-quarter sub-set. These are: (i) divide each segment into 4 and then, (ii) reading from L to R: delete the top 3/4 if the original segment is the 1st quarter, the top 1/2 if the original segment is 2nd quarter, and the top 1/4 if the original segment is 3rd quarter. There is no 4th quarter to consider because this is always deleted in the previous iteration. The resulting graphical construction of the fourth-quarter sub-set is shown (to the third iteration) in Fig. 5, superimposed on the corresponding construction of the full fourth-quarter set discussed earlier.

Finally, a ^1H NMR spectrum can be simulated purely from the numerical data shown in Table 1, with the code values treated as chemical shifts and the degeneracies as integrated intensities. The close agreement (Fig. 6 and inset) between this simulation and the experimental spectrum for the pyrene complex of copolymer **Z**

(Fig. 4) is very striking and provides good evidence that the “stop-at-zero” limitation is a real effect in such systems.

The question then arises of what the physical origins of the “stop-at-zero” limitation might be. An obvious clue is that spectra featuring this limitation are only seen for copolymers such as **Y** and **Z**, where pyrene is bound in the intercalative (dual-site) mode. Inspection of Table 1 shows that a key characteristic of the sequences emerging when the limitation is imposed is that the environment of the “central” 1 is defined—other than at the origin (code 000)—by a consecutive string of 1 s. In molecular terms, this corresponds to a consecutive run of NDI residues, linked in pairs by the chain-folding triethylene-dioxy residue “E” [15]. It is easy to see that, in such a sequence, the cumulative ring-current shielding of the central NDI residue by intercalating pyrene molecules would be disrupted by the presence of a non-binding comonomer unit “S”. Such a unit would tend to unfold the chain at that point, carrying any pyrene molecules subsequently bound at “E”-linked NDI pairs to locations much more distant from the central, “observed” NDI unit, where their magnetic shielding would then be negligible [15].

It might reasonably be asked whether the multiplication of NDI resonances observed on complexation of copolymers **X**, **Y** and **Z** with pyrene might arise (at least in part) from spin–spin (*J*-) coupling between inequivalent ^1H nuclei. This is, in principle, possible for the *ortho*-related protons of an NDI residue at the centre of an unsymmetrical sequence. However, we have reported in an earlier paper [17] that 2D-JRES analysis of the diimide resonances for copolymer **X**, in the presence of pyrene, shows no evidence whatever of *J*-coupling between the NDI protons, while *ortho*, *meta* and *para* couplings for the HFDI resonances are all readily identified. It is thus clear that *J*-coupling plays no part in generating the observed NDI resonance-patterns.

The experimental results cited in Refs. [15–17] were all derived for AB-type copolymers having a 1:1 molar ratio of the two comonomers and with an essentially random distribution of these within the copolymer chain. It might seem that these are special cases of AB-copolymers, and that the fractal nature of their NMR spectra could be eliminated in less random and/or non-equimolar copolymers. This, however, is not the case. The complexation shifts of NDI resonances in the presence of pyrene are determined *solely* by the distribution of other NDI residues in adjacent sequences, so that each sequence and each resonance is associated with just one specific shielding code (see for example Table 1). The identity of this code is unaffected by differences in copolymer stoichiometry or randomness of distribution, which factors—while directly influencing the relative *intensities* of the resonances [16]—do not change their shielding codes.

3 Conclusions

The intramolecular environment of a monomer residue within an AB-copolymer chain may be modelled in terms of an infinite string of ones and zeros. Summing digits at equivalent positions, in both directions, from any digit d of a given type (say $d=1$), affords a number—a *code*—which can in principle be in any base, b ,

higher than 2 (i.e. it cannot be a binary number). For bases higher than 3 the codes define a limited set of numbers comprising a last-fraction Smith–Cantor set. Experimentally, the ^1H NMR spectrum of a random, binary co(polyester-imide) shows, on complexation with the ring-current-shielding molecule pyrene, a pattern of chemical shifts approximating very closely to the fourth-quarter Smith–Cantor set (i.e. where $b=4$). This result indicates that, although the set of codes for such a copolymer has the potential to represent numbers in any base higher than 2, complexation with pyrene leads specifically to selection of the base 4. We interpret this result as indicating that the degree of magnetic shielding of a “central observed” NDI residue resulting from pyrene-complexation at a *neighbouring* NDI residue falls off exponentially—by a factor of approximately 4—as its numerical distance from the centre increases along the copolymer chain. While this premise leads to theoretical results in very close agreement with experiment, we have not presented any direct proof in support of the proposed exponential decay of the magnetic shielding. Computational simulations using, for example, density functional theory, might be able to confirm this point in the future. However, the complexity and dynamic character of the supramolecular systems involved make this a major undertaking, well beyond the scope of the present work. Specifically, the observed NMR spectra represent *time-averaged* results, not only of macromolecular chain-dynamics in solution but also of rapidly reversible binding (on the NMR time scale) of pyrene to the in-chain NDI residues, including the effects of such binding on the conformational characteristics of the copolymer chain itself.

Other co(polyimide) complexes show a different, but related, ^1H NMR pattern, now corresponding to a specific *sub-set* of the same fractal. For $d=1$, it is shown that this sub-set results from a “stop-at-zero” limitation, whereby digits in the initial two-digit string are disregarded (i.e. set to zero) for code-generating purposes if they occur beyond a zero, when viewed from the central “1”. This limitation is found to arise in copolymer systems where the shielding molecule binds by intercalation between *pairs* of adjacent NDI residues, leading to cumulative incremental shielding until a non-binding residue is reached, after which no additional shielding is observed.

Acknowledgements We wish to thank our many co-workers cited in Refs. [15–17], especially Dr John Shaw who was the first to discover evidence of fractality in the NMR spectra of copolymer complexes [15, 36]. Helpful guidance in the area of fractal mathematics was provided by Professor Kenneth Falconer and Dr Keith Hannabuss. The work was supported by the Engineering and Physical Research Council of the UK (Grant Numbers EP/E00413X/1 and EP/G026203/1), the European Union (Marie Skłodowska-Curie Network EURO-SEQUENCES, Grant Number 642083), the Leverhulme Foundation (an Emeritus Fellowship to HMC), and by the Universities of Reading and Cambridge.

Author contributions H.M.C. and R.G.-C. wrote the manuscript and H.M.C. prepared the figures. The manuscript was reviewed by both authors.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. F.H.C. Crick, On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163 (1958)
2. M.W. Nirenberg, J.H. Matthaei, The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* **47**, 1588–1602 (1961)
3. M. Nirenberg, Historical review: deciphering the genetic code—a personal account. *Trends Biochem. Sci.* **29**, 46–54 (2004)
4. C.R. Dawkins, *The Blind Watchmaker* (Longmans, London, 1986), p.115
5. H.M. Colquhoun, J.-F. Lutz, Information-containing macromolecules. *Nat. Chem.* **6**, 455–456 (2014)
6. M.G.T.A. Rutten, F.W. Vaandrager, J.A.A.W. Elemans, R.J.M. Nolte, Encoding information into polymers. *Nat. Rev. Chem.* **2**, 367–381 (2018)
7. J.-F. Lutz, M. Ouchi, D.R. Liu, M. Sawamoto, Sequence-controlled polymers. *Science* **341**, 1238149 (2013)
8. J.-F. Lutz, J.-M. Lehn, E.W. Meijer, K. Matyjaszewski, From precision polymers to complex materials and systems. *Nat. Rev. Chem.* **1**, 1–14 (2016)
9. D. Nunez-Villanueva, M. Ciaccia, G. Iadevaia, E. Sanna, C.A. Hunter, Sequence information transfer using covalent template-directed synthesis. *Chem. Sci.* **10**, 5258–5266 (2019)
10. F.T. Szczyplinski, L. Gabrielli, C.A. Hunter, Emergent supramolecular assembly properties of a recognition-encoded oligoester. *Chem. Sci.* **10**, 5397–5404 (2019)
11. L. Gabrielli, D. Nunez-Villanueva, C.A. Hunter, Two-component assembly of recognition-encoded oligomers that form stable H-bonded duplexes. *Chem. Sci.* **11**, 561–566 (2020)
12. H.M. Colquhoun, Z. Zhu, Recognition of polyimide sequence-information by a molecular tweezer. *Angew. Chem. Int. Ed.* **43**, 5040–5045 (2004)
13. H.M. Colquhoun, Z. Zhu, C.J. Cardin, Y. Gan, M.G.B. Drew, Sterically controlled recognition of macromolecular sequence information by molecular tweezers. *J. Am. Chem. Soc.* **129**, 16163–16174 (2007)
14. Z. Zhu, C.J. Cardin, Y. Gan, H.M. Colquhoun, Sequence-selective assembly of tweezer-molecules on linear templates enables frameshift reading of sequence information. *Nat. Chem.* **2**, 653–660 (2010)
15. J.S. Shaw, R. Vaiyapuri, M.P. Parker, C.A. Murray, K.J.C. Lim, C. Pan, M. Knappert, C.J. Cardin, B.W. Greenland, R. Grau-Crespo, H.M. Colquhoun, Elements of fractal geometry in the ^1H NMR spectrum of a copolymer intercalation-complex: identification of the underlying Cantor set. *Chem. Sci.* **9**, 4052–4061 (2018)
16. M. Knappert, T. Jin, S.D. Midgley, G. Wu, O.A. Scherman, R. Grau-Crespo, H.M. Colquhoun, Supramolecular complexation between chainfolding poly(ester-imide)s and polycyclic aromatics: a fractal-based pattern of NMR ring-current shielding. *Polym. Chem.* **10**, 6641–6650 (2019)
17. M. Knappert, T. Jin, S.D. Midgley, G. Wu, O.A. Scherman, R. Grau-Crespo, H.M. Colquhoun, Single-site binding of pyrene to poly(ester-imide)s incorporating long spacer-units: prediction of NMR resonance-patterns from a fractal model. *Chem. Sci.* **11**, 12165–12177 (2020)
18. Y. Ren, R. Jamagne, D.J. Tetlow, D.A. Leigh, A tape-reading molecular ratchet. *Nature* **612**, 78–82 (2022)
19. P. Lazzaretti, Ring currents. *Prog. Nucl. Magn. Reson. Spectrosc.* **36**, 1–88 (2000)

20. S. Klod and E. Kleinpeter, Ab initio calculation of the anisotropy effect of multiple bonds and the ring current effect of arenes—application in conformational and configurational analysis. *J. Chem. Soc., Perkin Trans. 2*, 1893–1898 (2001).
21. C.A. Hunter, J.K.M. Sanders, The nature of π – π interactions. *J. Am. Chem. Soc.* **112**, 5525–5534 (1990)
22. F. Cozzi, F. Ponzini, R. Annunziata, M. Cinquini, J. Siegel, Polar interactions between stacked π systems in fluorinated 1,8-diarylnaphthalenes: importance of quadrupole moments in molecular recognition. *Angew. Chem. Int. Ed.* **34**, 1019–1020 (1995)
23. J.F. Fleron, A note on the history of the Cantor set and Cantor function. *Math. Magn.* **67**, 136–140 (1994)
24. J.-L. Chabert, Un demi-siecle de fractales: 1870–1920. *Hist. Math.* **17**, 339–365 (1990)
25. H.J.S. Smith, On the integration of discontinuous functions. *Proc. Lond. Math. Soc.* **6**, 140–153 (1875)
26. G. Cantor, Grundlagen einer allgemeinen Mannigfaltigkeitslehre. *Math. Ann.* **21**, 545–591 (1883) (An English translation of this paper can be found at: <https://www.jamesrmeyer.com/infinite/cantor-grundlagen>)
27. K. Hannabuss, Forgotten fractals. *Math. Intell.* **18**, 28–31 (1996)
28. K. Hannabuss, in *Oxford's Savilian Professors of Geometry*, ed. by R.J. Wilson (Oxford University Press, Oxford, 2022), p. 93.
29. T. Hawkins, *Lebesgue's Theory of Integration: Its Origins and Development* (Chelsea Publishing, New York, 1975), pp.37–40
30. H.-O. Peitgen, H. Jürgens, D. Saupe, *Chaos and Fractals* (Springer, Berlin, 1992), pp.67–77
31. B.B. Mandelbrot, *The Fractal Geometry of Nature* (WH Freeman and Co., New York, 1982)
32. A. Rényi, *Acta Math. Acad. Sci. Hung.* **8**, 477–493 (1957)
33. Č Burdik, Ch. Frougny, J.P. Gazeau, R.J. Krejcar, *Physica A* **31**, 6449–6472 (1998)
34. B.L. Iverson, R.S. Lokey, *Nature* **375**, 303–305 (1995)
35. J.G. Hansen, N. Feeder, D.G. Hamilton, M.J. Gunter, J. Becher, J.K.M. Sanders, *Org. Lett.* **2**, 449–452 (2000)
36. J.S. Shaw, PhD Thesis (University of Reading, 2011)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.