doi.org/10.26434/chemrxiv.11860881.v1

Sequence-Modification in Copoly(ester-Imide)s: A Catalytic/supramolecular Approach to Writing/reading Copolymer Sequence-Information

Marcus Knappert, Howard Colquhoun

Submitted date: 17/02/2020 • Posted date: 19/02/2020

Citation information: Knappert, Marcus; Colquhoun, Howard (2020): Sequence-Modification in Copoly(ester-Imide)s: A Catalytic/supramolecular Approach to Writing/reading Copolymer Sequence-Information. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11860881.v1

Catalytic ester-interchange reactions, analogous to mutation and recombination, allow new sequence-information to be written, statistically, into NDI-based poly(ester-imide) chains. Thus, insertion of the cyclic ester cyclopentadecanolide ("exaltolide") into an NDI-based homopolymer, and quantitative sequence-exchange between two different homopoly(ester-imide)s, are catalysed by di-n-butyl tin(IV) oxide. Emerging sequences are identified at the triplet and quintet levels by ^{1}H NMR analysis, using supramolecular complexation of pyrene-d_{10} at the NDI residues to amplify the separation of resonances associated with different sequences. In such systems, pyrene is able to act as a "reader-molecule" by generating different levels of ring-current shielding from the different patterns of supramolecular binding to all the NDI-centred sequences of a given length.

File list (2)

Sequence-modification in copoly(ester-imide)s: a catalytic/supramolecular approach to writing/reading copolymer sequence-information

Marcus Knappert and Howard M. Colquhoun*

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

ABSTRACT: Catalytic ester-interchange reactions, analogous to mutation and recombination, allow new sequence-information to be written, statistically, into NDI-based poly(ester-imide) chains. Thus, insertion of the cyclic ester cyclopentadecanolide ("exaltolide") into an NDI-based homopolymer, and quantitative sequence-exchange between two different homopoly(ester-imide)s, are catalysed by di-*n*-butyl tin(IV) oxide. Emerging sequences are identified at the triplet and quintet levels by ^{1}H NMR analysis, using supramolecular complexation of pyrene-d_{10} at the NDI residues to amplify the separation of resonances associated with different sequences. In such systems, pyrene is able to act as a "reader-molecule" by generating different levels of ring-current shielding from the different patterns of supramolecular binding to all the NDI-centred sequences of a given length.

* Email: h.m.colquhoun@rdg.ac.uk

1. Introduction

The processing of information in biological systems involves a vast and complex array of molecular machinery,[1] most of which is devoted to "reading" copolymer sequence information,[2] notably through the transcription (from DNA to RNA)[3] and translation (from RNA to protein)[4] of genetically-relevant sections of polynucleotide chains. Highly specific editing of transcribed *m*-RNA sequences by enzymic cutting and splicing may also occur before translation, to remove non-coding sequences from the transcript.[5] However, modification of existing *DNA* sequences (i.e. the "writing" of new DNA sequence-information) is very much less specific and occurs mainly by statistical mechanisms such as mutation[6] and genetic recombination ("crossing-over").[7] Beneficial modifications of DNA are retained, and damaging ones discarded, by natural selection at the organism level. Recombination involves a limited exchange of sequences between closely-related segments of DNA chains originating in different parental genomes; mutation can involve the random insertion, or deletion of just a single nucleotide; but in chemical terms, both mutation and recombination are just interchange reactions at phosphodiester linkages in copolymer chains. Catalytic ester-interchange is a well-established reaction in polymer chemistry,[8,9,10] and so, in the present work, we sought to develop synthetic analogues of mutation and recombination, complementing earlier work on "reading" comonomer sequences in copoly(ester-imide)s.[11-14]

It has been reported that supramolecular complexation of a polycyclic aromatic molecule such as pyrene to diimide residues in binary copolyimides, via complementary π-π-stacking,[15] produces highly sequence-dependent NMR complexation shifts of the diimide resonances.[13,14] This results from cumulative ring-current shielding[16] by the pyrene "reader-molecules" as they bind at different sets of positions in different sequences, and has allowed the resolution and assignment of diimide ^{1}H resonances to sequences containing up to seven diimide residues.[13] Most recently, analogous results were observed (at the quintet level) for

copoly(*ester*-imide)s.[14] It then seemed to us possible that new sequence-information might be "written" into such polymers by catalytic ester-interchange, with the ring-current shielding method allowing new sequences to be "read" as they evolve. Here we report the realisation of this catalytic/supramolecular approach to the evolution of copolymer sequence-information.

2. Experimental Section

2.1 Materials and Instrumentation

Starting materials, monomers, solvents and analytical instrumentation were as described in recent publications.[13,14] Di-*n*-butyl tin(IV) oxide and cyclopentadecanolide ("exaltolide") were obtained from Sigma-Aldrich and used as received. Homopolymers **1**, **2** and **3** and copolymer **5** were synthesised and characterised as previously reported.[13,14]

2.1 Insertion of exaltolide into homopolymer 1

The NDI-based homopolymer **1** (0.421 g, 0.903 mmol per repeat unit), exaltolide (0.117 g, 0.488 mmol), di-*n*-butyl tin(IV) oxide (0.027 g, 0.108 mmol) and 1-chloronaphthalene (1.2 mL) were heated in a round-bottomed flask, under nitrogen, to 180 °C. The resulting homogeneous solution was held at this temperature for 4 hours. Aliquots (0.3 mL) were removed after 1, 2, and 3 hours and dissolved in chloroform/trifluoroethanol (6:1 v/v, 4 mL). The resulting solutions were each added with stirring to cold methanol (100 mL) and the polymer was filtered off, washed with methanol and dried at 60 °C under vacuum. A ^{1}H NMR spectrum of each copolymer sample was run in $CDCl_3$/TFE (6:1 v/v) and a second spectrum of each was taken in the presence of 5 equivalents per NDI residue of pyrene-d_{10}.

2.2 Sequence-exchange between homopolymers 2 and 3

The NDI-based homopolymer **2** (1.386 g, 2.880 mmol per repeat unit), the FDI-based homopolymer **3** (1.902 g, 2.840 mmol per repeat unit), di-*n*-butyl tin(IV) oxide (0.063 g,

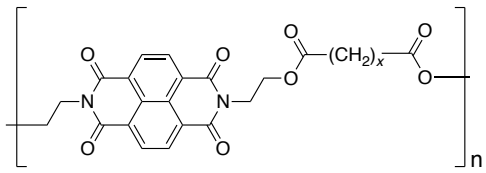
0.145 mmol) and 1,2-dichlorobenzene (7 mL) were ground together in a mortar and the resulting suspension was transferred to a round-bottomed flask and heated to reflux (180 °C) with stirring under a nitrogen atmosphere for 4 hours. Aliquots (0.5 mL) of reaction solution were removed at ten minute intervals for the first hour, under a counterflow of nitrogen, and diluted with dichloromethane/trifluoroethanol (3:1 v:v, 1 mL). When the aliquot had fully dissolved, the solution was added with stirring to cold methanol (100 mL) and the precipitated polymer was filtered off, washed with methanol and dried at 60 °C under vacuum. A fresh charge of di-*n*-buyl tin(IV) oxide (0.063 g) was added to the system at hourly intervals to compensate for catalyst deactivation, and a final polymer sample was taken after 4 h, when the reaction was found to be complete. A ^{1}H NMR spectrum of each sample was run in $CDCl_3$/TFE (6:1 v/v) and a second spectrum of each was taken in the presence of 7 equivalents per NDI residue of pyrene-d_{10}.

3. Results and Discussion

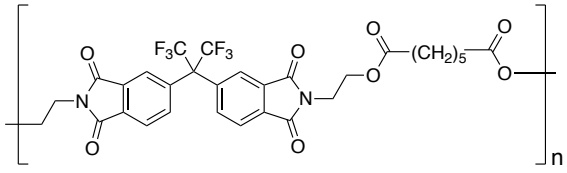
3.1 Insertion chemistry

The homopoly(ester-imide) **1** (Chart 1), formed by high-temperature polycondensation of pentanedioyl (glutaryl) chloride with the NDI-based diol **6**,[13] was found to react with the cyclic ester exaltolide (**7**) at 180 °C, in 1-chloronaphthalene as solvent, in the presence of di-*n*-butyl tin oxide as catalyst. A 2:1 molar ratio of repeat units between **1** and **6** was used, and the catalyst concentration was 5 wt% on reactants. Analysis by ^{1}H NMR spectroscopy showed that, after an hour, the single NDI resonance present in the spectrum of the starting homopolymer **1** had been replaced by a group of three closely-spaced diimide resonances with initially-estimated integrals of 2:5:3 (high to low field, Figure 1b). Addition of pyrene-d_{10} (5 equivalents per NDI residue) to the 1-hour sample resulted in upfield shifts and increased separation of the three NDI resonances (Figure 1c), allowing better definition of

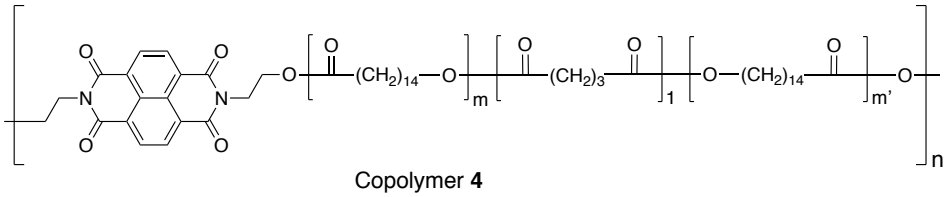
their relative integrals, although the presence of solvent and trace by-product resonances limited the accuracy of integration. Samples of the product at more extended reaction times showed no further change in the ^{1}H NMR spectrum, indicating that the insertion reaction was complete in less than an hour.



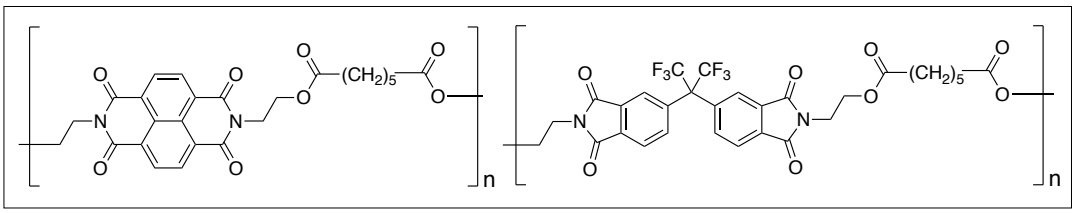
Homopolymers **1** and **2** (x = 3 and 5)



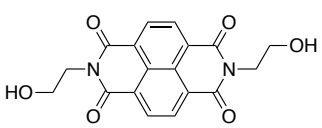
Homopolymer **3**



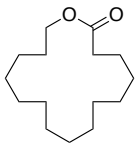
Copolymer **4**



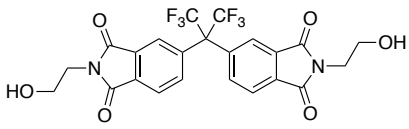
Copolymer **5**



6



7



8

Chart 1. Homopolymers, copolymers and starting materials used in the present work

A theoretical analysis (see ESI) of the potential for insertion of lactone **7** (exaltolide) into homopolymer **1** shows that (i) reaction can occur at one or both ester linkages between the glutaric acid and diimide-diol residues, (ii) single or multiple exaltolide residues can insert at each of these positions, (iii) because of the symmetrical (diacid) nature of the glutaryl unit, the direction of exaltolide insertion must be reversed about the glutaryl residue (see Chart1),

and (iv) for the same reason, only a single glutaryl residue can be present between any two successive NDI residues in the final copolymer (**4**). The structure of **4** can thus be abbreviated as $[\text{-I}(\text{E}_m)\text{G}(\text{E}_{m'})\text{-}]_n$, where I = NDI-diol, E = exaltolide, and G = glutaryl residues. The numbers of exaltolide repeat units *m* and *m′* in **4** can take any value, including zero. If both are zero then we simply have the starting material, $[\text{-I-G-}]_n$, homopolymer **1**. Consequently, the formulation ($\text{E}_m\text{GE}_{m'}$) describes *all possible* linking groups between any two **I** residues in both homopolymer **1** and copolymer **4**.

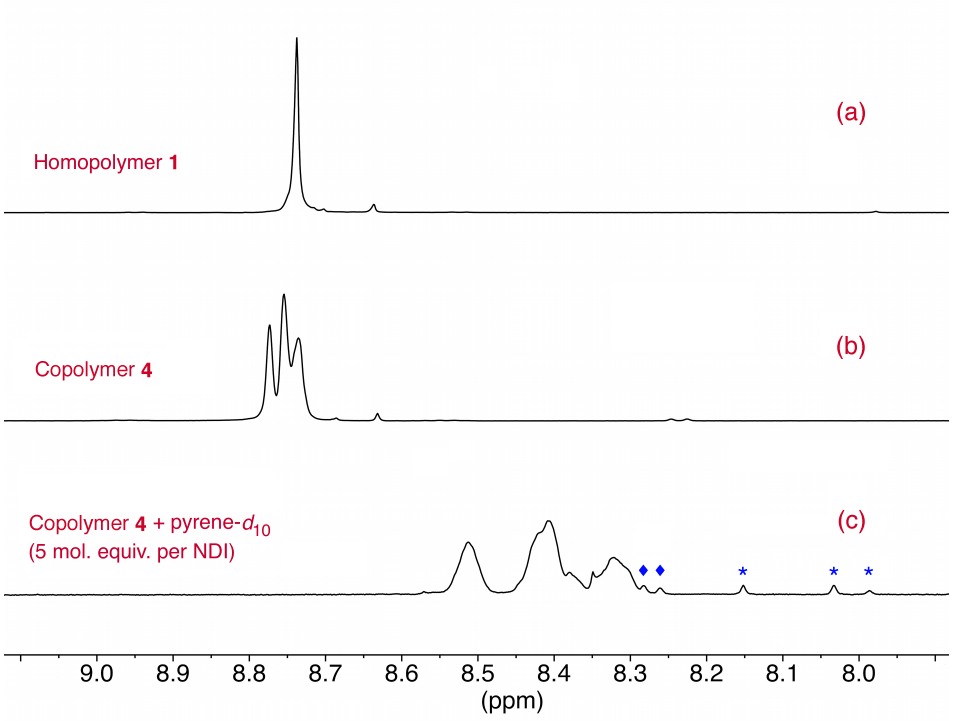


Figure 1. ^{1}H NMR spectra, in the NDI region, of (a) homopolymer **1**; (b) copolymer **4**, formed by insertion of exaltolide (**7**) into homopolymer **1**; and (c) copolymer **4** in the presence of pyrene-d_{10} (5 mol equiv. per NDI residue). Stars indicate resonances from residual protons in pyrene-d_{10}, and diamonds indicate resonances from residual reaction-solvent (1-chloronaphthalene). These and other impurity peaks limit the accuracy of integration in spectrum 1c, but relative values for the three principal resonances in spectrum 1c are estimated as 2.0 : 5.1 : 2.4 (high to low field).

The simplest interpretation of the three NDI resonances seen in the ^{1}H NMR spectrum of copolymer **4** is that they represent different copolymer sequences derived from the starting

"three-I" sequence [-IGIGI-]. As discussed in earlier papers, only the central diimide resonance need to be considered in this type of analysis, as the outer diimide residues are at the centres of other, analogous, sequences and so can be treated separately. After insertion of exaltolide, a "central" diimide residue in copolymer **4** may thus be present in sequences [-IGIGI-], [-I($E_mGE_{m'}$)IGI-], [-IGI($E_mGE_{m'}$)I-] and [-I($E_mGE_{m'}$)I($E_mGE_{m'}$)I-]. The second and third sequences here are directionally degenerate and so would not be distinguishable in the ^{1}H NMR spectrum, leaving just three potential NDI resonances for these four sequences.

The general linking group ($E_mGE_{m'}$) between successive NDI residues can of course represent many different chemical entities, depending on the values of m and m', but as "E" is a linear, sixteen-atom moiety, even a single exaltolide insertion would more than double the number of atoms between successive I residues. Consequently, as far as the central NDI residue is concerned, the only significant distinguishing feature between its different environments is whether there are zero, one, or two "E" residues present in the units linking it to its neighbouring NDIs. Thus it only remains to calculate the probabilities of the sequences [-IGIGI-], [-I($E_mGE_{m'}$)IGI-]/[-IGI($E_mGE_{m'}$)I-] and [-I($E_mGE_{m'}$)I($E_mGE_{m'}$)I-] to give a prediction of the relative intensities (i.e. integrals) of the corresponding ^{1}H NMR resonances.

It can be shown (see ESI) that, in copolymer **4**, the probability of a linker between two **I**'s being simply G (i.e. m = m' = 0) is 0.667, and that the probability of the linker being ($E_mGE_{m'}$) for all other values of m or m', taken together, is 0.833. The probability of an I, rather than an E being found at the end of each sequence is also 0.667, but the probability of the central I must, by definition (as we are considering only I-centred sequences), be 1. Multiplying up the individual probabilities to give values for each sequence, we then have:

Sequence	Residue probabilities		Sequence Probability
IGIGI	1 x $(0.667)^4$	=	0.198
I($E_nGE_{n'}$)IGI	1 x (0.833) x $(0.667)^3$	=	0.247
IGI($E_nGE_{n'}$)I	1 x (0.833) x $(0.666)^3$	=	0.247
			0.494
I($E_nGE_{n'}$)I($E_nGE_{n'}$)I	1 x $(0.833)^2$ x $(0.667)^2$	=	0.308
	Total probability	=	1.000

Since the total probability, as shown, equals 1, this analysis must include all the probabilities for all the sequences in copolymer **4** that are based on three successive NDI residues. The relative intensities of NDI resonances in copolymer **4** are thus predicted to be 0.198 : 0.494 : 0.308. These values are reasonably close to those found experimentally in the ^{1}H NMR spectrum of the insertion product, copolymer **4**, confirming that an effective sequence-modification reaction analogous to mutation has been achieved, and providing assignments of the three observed resonances to the sequences [-IGIGI-], [-I($E_mGE_{m'}$)IGI-]/[-IGI($E_mGE_{m'}$)I-] and [-I($E_mGE_{m'}$)I($E_mGE_{m'}$)I-] (high to low field respectively).

3.2 Sequence-exchange chemistry

Equimolar proportions (in diimide residues) of the homopoly(ester-imide)s **2** and **3** (Chart 1), were heated at 180 °C under nitrogen in 1,2-dichlorobenzene as solvent, in the presence of di-*n*-butyl tin(IV) oxide as catalyst. Samples of the reaction mixture were taken at 10-minute intervals and the polymeric component was isolated and analysed by ^{1}H NMR spectroscopy. Initially, the ^{1}H NMR spectrum of a mixture of the two starting homopolymers showed just a single NDI resonance: addition of pyrene-d_{10} (7 equivalents per NDI) to the NMR sample shifted this signal upfield by ca. 0.5 ppm, but no splitting of the resonance was observed (Figure 2). However, after a reaction time of one hour, again using pyrene addition to expand the spectrum, new resonances appeared at lower field (Figure 2) and, from previous work,

these could be assigned to sequences containing both NDI and HFDI residues.[14] After 2 hours, further new resonances were evident at still lower field, and the highest-field (all-NDI) resonance had diminished substantially in intensity. Finally, after four hours, the original (single) NDI peak had been replaced by no fewer than nine NDI resonances in the range 8.25 – 8.50 ppm (spectrum measured in the presence of 7 equivalents of pyrene-d_{10} per NDI), with approximate relative intensities of 1:2:1:2:4:2:1:2:1 (Figure 2). This final pattern was previously identified as being characteristic of the fully-random, binary copolymer **5**, synthesised by statistical co-polycondensation of monomers **6** and **8** (equimolar ratio) with heptanedioyl (pimeloyl) chloride.[14]

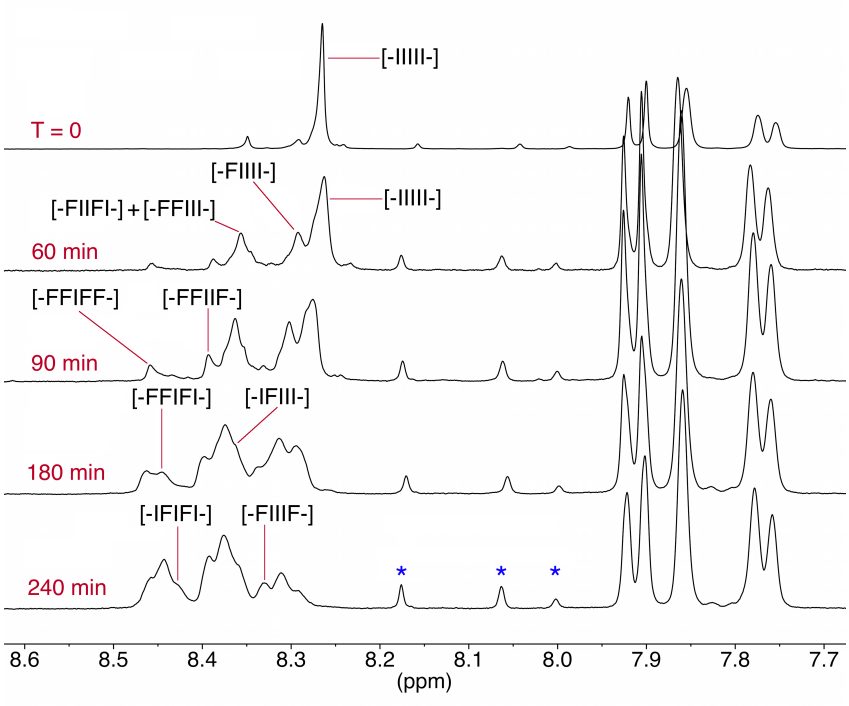


Figure 2. ^{1}H NMR spectra showing the progressive development of new quintet sequences resulting from tin-catalysed interchange of sequences between homopolymers **2** and **3**, a process leading ultimately (240 min) to the formation of random copolymer **5**. "I" = NDI and "F" = HFDI. Peaks between 7.70 7.95 ppm are due to HFDI protons and the starred resonances between 8.00 and 8.20 ppm represent residual protons in pyrene-d_{10} (7 equivalents added per mole of NDI residues). For peak assignments see ref. 14, Table 1).

An earlier analysis of this nine-line pattern, based on single-site binding of pyrene to each NDI ("I") residue in the copolymer, showed that it arises from a fractal distribution of ring-current shieldings that sum to different values depending on the sequence involved.[14]

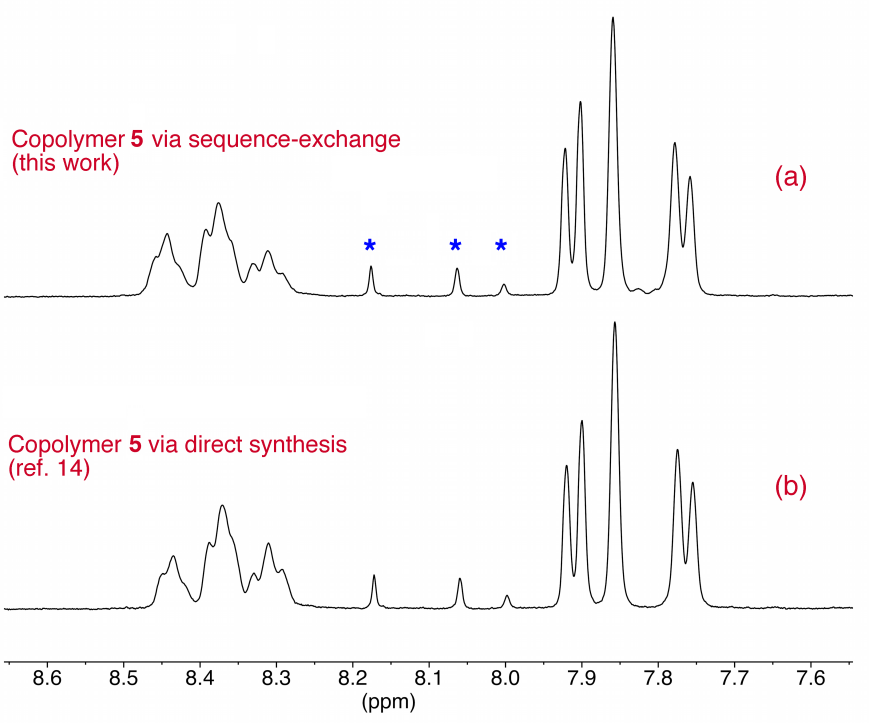


Figure 3. ^{1}H NMR spectra of copolymer **5** in the presence of pyrene-d_{10}. (a) sample from ester-interchange (this work), and (b) sample from direct synthesis.[14] Peaks between 7.70 and 7.95 ppm are due to HFDI protons and the starred resonances between 8.00 and 8.20 ppm represent residual protons in (99.8% D) pyrene-d_{10}.

This analysis allowed all nine resonances to be assigned to specific NDI-centred quintets (or groups of quintets) so that here, as shown in Figure 2, we can follow in detail the emergence of new sequences from ester-interchange between the starting homopolymers **2** and **3**. Finally, the ^{1}H NMR spectra of samples of copolymer **5** obtained by (a) direct copolymerisation[14] and (b) sequence-exchange are compared in Figure 3, using pyrene-d_{10} to resolve sequence-information at the quintet level. It is evident from this result that catalytic sequence-exchange can indeed provide a successful approach to the statistical "writing" of new sequences in poly(ester-imide) systems.

4. Conclusions

Catalytic ester-interchange reactions, analogous to mutation and recombination, enable new sequence-information to be written, statistically, into NDI-based poly(ester-imide) chains. Emerging sequences may be identified at the triplet and quintet levels by ^{1}H NMR analysis using supramolecular complexation of pyrene-d_{10} to the diimide residues. In such systems, pyrene acts as a "reader-molecule" by generating different levels of ring-current shielding from the different patterns of binding to all possible NDI sequences of any given length.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

This work was sponsored by the H2020 programme of the European Union under the ITN project *Euro-Sequences*, H2020-MSCA-ITN-2014, grant number 642083 (Marie Skłodowska-Curie PhD studentship to MK). The final stages of the work were supported by the Leverhulme Trust (Emeritus Fellowship to HMC, grant number EM-2018-0161/4). We are grateful to Dr Ricardo Grau-Crespo of the University of Reading for valuable discussions.

References

1. *The Machinery of Life*, D.S. Goodsell, Springer, New York, 2nd Ed., 2009.

2. C. Adami, Information theory in molecular biology, *Phys. Life Rev*., 2004, **1**, 3–22.

3. D. A. Bushnell, K. D. Westover, R. E. Davis and R. D. Kornberg, Structural basis of transcription: an RNA polymerase II–TFIIB cocrystal at 4.5 angstroms. *Science*, 2004, **303**, 983–988.

4. C. O. Gualerzi and C. L. Pon, Initiation of mRNA translation in prokaryotes. *Biochemistry*, 1990, **29**, 5881–5889.

5. T. Maniatis, "Split genes", in *Encyclopedia of Genetics*, S. Brenner and G. H. Miller (eds.), Academic Press, New York, 2001, p. 1878.

6. General nature of the genetic code for proteins, F. H. C. Crick, L. Barnett, S. Brenner and R. J. Watts-Tobin. *Nature*, 1961, **192**, 1227–1232.

7. R. Holliday, Recombination and meiosis. *Philos. Trans. R. Soc. B*, 1977, **277** (955), 359–370

8. M. Shuster, M. Narkis and A. Siegmann, Catalytic transesterification in polycarbonate-polycaprolactone systems. *J. Appl. Polym. Sci*. 1994, **52**, 1994, 1383–1391.

9. P. Hodge, Z. Yang, A. Ben-Haida and C. S. McGrail, Cyclo-depolymerisation of poly (ethylene naphthalene-2,6-dicarboxylate) and ring-opening polymerisations of the cyclic oligomers obtained. *J. Mater. Chem*., 2000, **10**, 1533–1537.

10. N. González-Vidal, A. Martínez de Ilarduya, V. Herrera and S. Muñoz-Guerra, Poly(hexamethylene terephthalate-co-caprolactone) copolyesters obtained by ring-opening polymerization. *Macromolecules*, 2008, **41**, 4136–4146.

11. Z. Zhu, C. J. Cardin, Y. Gan and H. M. Colquhoun, Sequence-selective assembly of tweezer-molecules on linear templates enables frameshift reading of sequence information. *Nat. Chem*., 2010, **2**, 653–660.

12. 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 and H. M. Colquhoun, Elements of fractal geometry in the ^{1}H NMR spectrum of a copolymer intercalation-complex: identification of the underlying Cantor set. *Chem. Sci*., 2018, **9**, 4052–4061.

13. M. Knappert, T. Jin, S. D. Midgley, G. Wu, O. A. Scherman, R. Grau-Crespo and 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*. 2019, **10**, 6641–6650.

14. M. Knappert, T. Jin, S. D. Midgley, G. Wu, O. A. Scherman, R. Grau-Crespo and 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. *ChemRxiv* Preprint, 2020. https://doi.org/10.26434/chemrxiv.11836251.v2.

15. C. A. Hunter and J. K. M. Sanders, The nature of π-π interactions. *J. Am. Chem. Soc*., 1990, **112**, 5525–5534.

16. 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*, 2001, 1893–1898.

Sequence-modification in copoly(ester-imide)s: a catalytic/supramolecular approach to writing and reading copolymer sequence-information

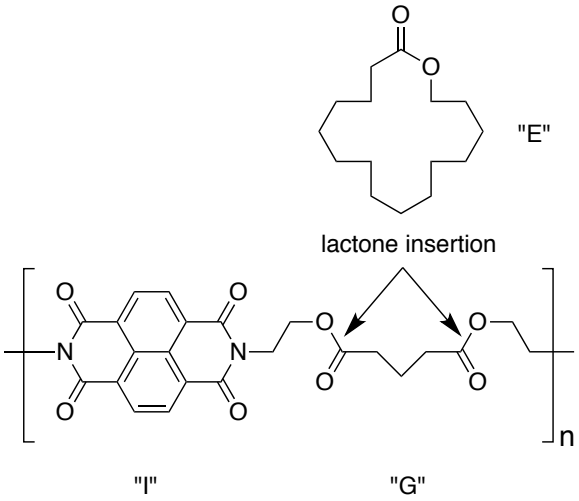
Marcus Knappert and Howard M. Colquhoun*

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

Electronic Supporting Information

Statistical insertion theory

The three monomer residues (ND**I**-diol, **G**lutaric acid and **E**xaltolide), present in the copoly(ester-imide) (**4**) formed by insertion of exaltolide into homopolymer **1**, are given the abbreviations I, G, and E respectively, as shown below. From the reaction stoichiometry (one mole of exaltolide to two moles of NDI), I, G and E should be present in the ratio 2:2:1 in the final copolymer.



In terms of the inserted "E" residues, there are only three options *for the immediate environment of the central* **I** *residue* in a "triplet" sequence I-----**I**-----I:

- No E on either side,
- One E (or more) on one side or the other.
- One E (or more) on both sides

* Email: h.m.colquhoun@rdg.ac.uk

The ability of "E" to propagate, giving a consecutive series of E's by repeated insertion, is accounted for as described below:

1. Starting from the central NDI-diol residue "I", calculate out in both directions. The probability of the central I is **one**, by definition.

2. The first-adjacent residue to I can be either G or E, but not I because I has no COOH group. The ratio of G and E in the system is 2:1, so the probability of the *first*-adjacent residue being G is 0.667 and the probability of it being E is 0.333.

3. If the first-adjacent residue to I is E, then this must have a hydroxyl OH end-group, which means it can link either with another E, through the latter's COOH group, or with a G, but not with an I (because this has no COOH group). The ratio of G and E in the system is 2:1, so the probability of the *second*-adjacent residue (following an E) being G is again 0.667 and the probability of it being E is again 0.333.

4. If the first-adjacent residue is G, then this must have a COOH end-group, so it can link with either E (though its hydroxyl OH group) or I, but not with another G (because G has no hydroxyl OH group). The molar ratio of I to E in the system is 2:1, so the probability of the *second*-adjacent residue (following a G) being I is again 0.667 and of it being E is again 0.333.

5. More generally, we can define the linkers from the central I to terminal I-groups as ($E_mGE_{m'}$). The initial (say left-side) E must link to the central I through its carboxyl COOH group, meaning that the other end (and all ends for further inserted E_n) must be hydroxyl. When these ends link to G, the symmetrical nature of G means that the direction of the E-repeats must *reverse*, so that the terminal group is then always carboxyl COOH. The latter cannot link to another G, only to a hydroxyl OH group of I, or to another E. So ($E_mGE_{m'}$) always represents a dicarboxylic acid residue and is therefore a *complete description of all possible linkers* between I residues in the copolymer product.

6. If we now define the sequences of interest as only those containing *three* I residues, then the possible linkers between the I's are just G or ($E_mGE_{m'}$). Of course G is simply ($E_nGE_{n'}$) where both *m* and *m′* are zero. The effects on the NMR spectrum are proposed to be that, when the linker is G, pyrene-binding to a neighbouring I is close

enough to produce an additional complexation shift at the "observed" I, but when one or more E's are inserted between two I's, they are too widely spaced for there to be any additional "long-range" shielding effects. The linker "G" itself is therefore treated separately, so that the possible squences within a framework I-----I-----I are:

IGIGI: I($E_mGE_{m'}$)IGI and IGI($E_mGE_{m'}$)I: I($E_mGE_{m'}$)I($E_mGE_{m'}$)I:

7. We can now have only two possible linkers, G and ($E_mGE_{m'}$), so we next need to calculate the relative probabilities of these.

Analysing the allowed diacid linkers between any two I residues we have:

G (HOOC-C_5-COOH) (Length 1)

EG (HOOC-C_{14}-OOC-C_5-COOH) and (Length 2)
GE (HOOC-C_5-COO-C_{14}-COOH (Length 2)

EEG (HOOC-C_{14}-OOC-C_{14}-OOC-C_5-COOH) and (Length 3)
GEE (HOOC-C_5-OOC-C_{14}-OOC-C_{14}-COOH) and (Length 3)
EGE (HOOC-C_{14}-OOC-C_5-COO-C_{14}-COOH) (Length 3)

Continuing, but omitting chemical detail, we have:

EEEG, GEEE, EEGE, EGEE; (Length 4)

EEEEG, GEEEE, EGEEE, EEEGE, EEGEE. (Length 5)

EEEEEG, EEEEEG, EGEEEE, EEEEGE, EEGEEE, EEEGEE (Length 6)

The emerging pattern is that the *number* of linkers of a given length equals the *length* of the linker. We can only have one G per linking sequence, because the first G causes propagation of E's in both directions so as to leave an "E" carboxyl COOH group at each end. These cannot link with G (which also has only carboxyl groups), so there can only be one G per linking sequence between any two I's.

8. The mole ratio of G to E is 2:1, so the probability of G (the shortest possible linker) is 0.667, and the probability of E is 0.333. The probability of linker GE is then 0.667 x 0.333 = 0.222. EGE would be 0.074, and EGEE = 0.025, etc. But this takes no account of sequence degeneracy. We have shown (above) that each ($E_nGE_{n'}$) sequence has a degeneracy equal to the length of the linker, so to get the total probability we need to multiply each linker-probability by the corresponding linker length (i.e. the number of E and G residues it contains).

Length 1	0.667 x 1	=	0.667 ("G")
Length 2	0.222 x 2	=	0.444
Length 3	0.074 x 3	=	0.222
Length 4	0.025 x 4	=	0.100
Length 5	0.008 x 5	=	0.040
Length 6	0.003 x 6	=	0.018
Sum of probabilities		=	1.490, i.e. converging to a limit of 1.5

Separating G out from all the other ($E_nGE_{n'}$)'s then gives linker-probabilities of 0.667 (G) and 0.833 ($E_mGE_{m'}$) (total = 1.5). But these refer to sequences *without* the terminal I-groups, and we need to include the probability of the I's at end of each sequence. The ratio of I to E is 2:1, so the probability of an E (carboxyl-ended) being followed by I is 0.667 and by E is 0.333. If the linker ends in G, then again the probability of the next group being I is 0.667 and of it being E is 0.333.

Thus we have:

Sequence	**Residue probabilities**		**Sequence Probability (= resonance intensity)**
IGIGI	1 x $(0.667)^4$	=	0.198
I($E_mGE_{m'}$)IGI	1 x (0.833) x $(0.667)^3$	=	0.247
IGI($E_mGE_{m'}$)I	1 x $(0.667)^3$ x (0.833)	=	0.247
			0.494
I($E_mGE_{m'}$)I($E_mGE_{m'}$)I	1 x $(0.833)^2$ x $(0.667)^2$	=	0.308
	Total probability	=	1.000