Molecular Recognition

Recognition of Polyimide Sequence Information by a Molecular Tweezer**

Howard M. Colquhoun and Zhixue Zhu*

The recognition of sequence information in linear macromolecules underpins the whole of biology, most notably perhaps in protein synthesis through the operation of the genetic code,[1] but also in the functioning of genetic regulatory proteins,[2] of restriction enzymes,[3] and of drug molecules that bind to specific nucleotide sequences in DNA.[4] However, nucleic acids, when abstracted from their biological context, are simply random, four-monomer copolymers with no inherent order or information content. Their information-bearing function derives solely from the ability of other molecules to recognize specific monomer sequences through noncovalent interactions. The recognition of sequence information in high-molecular-weight copolymers by molecules of complementary structure is thus an essentially biological phenomenon.[5] Sequence recognition in small (typically trimeric) peptides has, however, been achieved by using artificial "tweezer-type" peptides as receptors.[6–11] Progress has also been made in the search for artificial enzymes by combinatorial attachment of amide side

[*] Prof. H. M. Colquhoun, Dr. Z. Zhu
School of Chemistry
University of Reading
Whiteknights, Reading, RG6 6AD (UK)
Fax: (+44) 118-378-8450
E-mail: h.m.colquhoun@rdg.ac.uk

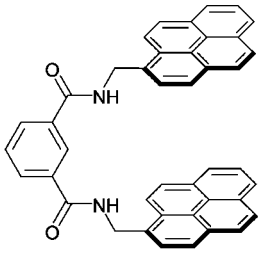
[**] This work was supported by the University of Reading Research Endowment Trust and by the DuPont Corporation (European University Research Grant).

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

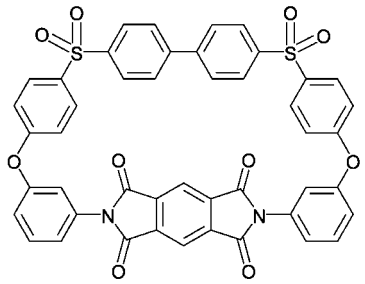
  DOI: 10.1002/ange.200460382

chains to polyamines, but the sequences responsible for catalytic activity in these latter systems are entirely unknown.[12] Future technologies involving the storage and transcription of information on the molecular scale will certainly require materials with far greater stabilities than those of peptides and nucleic acids. Herein we investigate the possibility of supramolecular sequence recognition in high-molecular-weight aromatic polyimides, a class of materials with outstanding resistance to thermochemical degradation.[13] We have discovered that extended yet completely specific monomer sequences in copolyimides can be recognized through their interactions with a sterically and electronically complementary molecular tweezer.

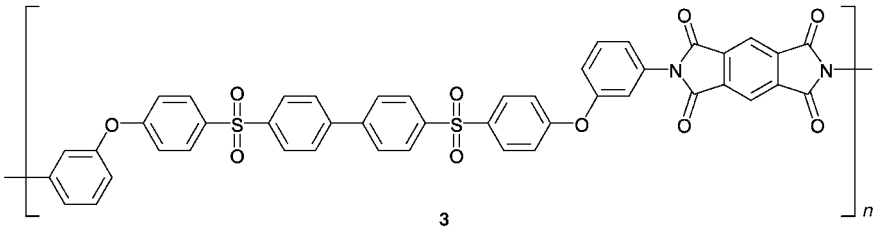
We have shown previously that the molecular tweezer **1** binds strongly to pyromellitimide residues through complementary π–π stacking and hydrogen-bonding interactions and



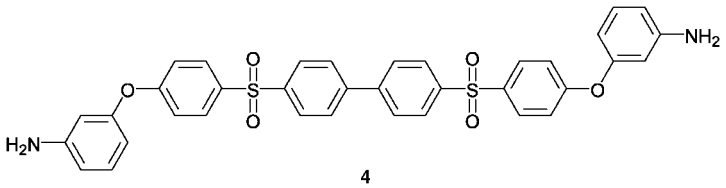
1



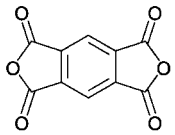
2



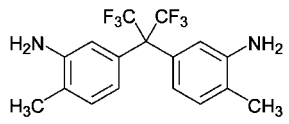
3



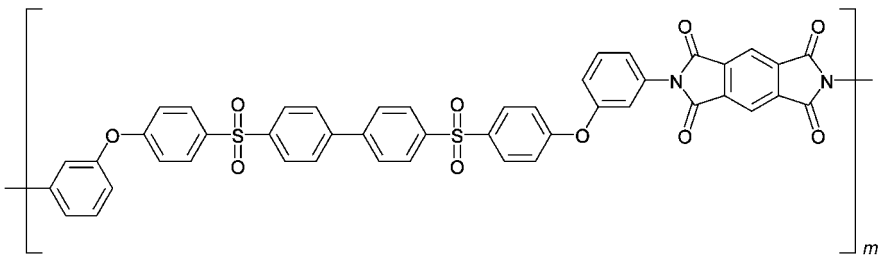
4



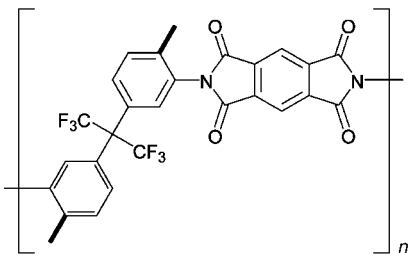
5



6



7 ($m = 2.33n$)



8 ($m = n$)

that, in complexes with cyclic oligomers such as **2**, tweezer binding can be enhanced through further π-stacking with the adjacent 4,4′-biphenylenedisulfone unit.[14] Extrapolation from the X-ray crystal structure of one such complex through computational modeling studies[15] suggested that a set of analogous interactions should also occur between tweezer **1** and polymer **3** (the linear, high-molecular-weight homologue of macrocycle **2**). Specifically, these studies predicted that the binding of **1** to a pyromellitimide residue in the polymer chain would be promoted by chain folding of adjacent ether–sulfone units at the 3-aminophenoxy residue so bringing their 4,4′-biphenylenedisulfone residues into π-stacking contact with the pyrenyl arms of the tweezer (Figure 1 a).

The molecular tweezer **1** was obtained by condensation of isophthaloyl chloride with 1-pyrenemethylamine,[14] and high-molecular-weight polyimide **3** ($M_n \approx 160 \times 10^3$ Da by GPC) was synthesized by polycondensation of diamine **4** with pyromellitic dianhydride (**5**) (see Supporting Information). Addition of tweezer **1** to a solution of polyimide **3** in chloroform/hexafluoropropan-2-ol gave an immediate deep-red solution, which was attributed to the charge-transfer absorption resulting from imide–tweezer π-stacking interactions. In keeping with computational predictions, ^{1}H NMR spectroscopy studies across a wide range of tweezer-to-imide molar ratios (from 1:80 to 4:1) confirmed that both imide and 4,4′-biphenylenesulfone residues are involved in binding to tweezer **1**. In the presence of the tweezer, large aromatic ring-current shifts were observed for the pyromellitimide proton resonance ($\Delta\delta = 2.6$ ppm at 4:1 tweezer/imide mole ratio) and for the 4,4′-biphenylene resonances ($\Delta\delta = 0.8$ ppm for the protons *ortho* to the biaryl linkage; protons *meta* to this linkage lie close to the boundary between shielding and deshielding zones of the pyrene tweezer arm and are thus only slightly affected by complexation). However, no significant evidence for tweezer binding was found for non-imide-based polymers containing the 4,4′-biphenylenedisulfone residue, thus confirming that the pyromellitimide unit in polymer **3** represents the primary tweezer-binding site, and that neighboring 4,4′-biphenylenedisulfone units provide only secondary stabilizing interactions. Despite the high binding constant for the complexation of tweezer **1** with polymer **3** ($K_a = 8.6(\pm 0.9) \times 10^3\,\text{M}^{-1}$), which was determined from the charge-transfer absorption by a UV/Vis dilution technique,[16] the pyromellitimide resonance remains as a sharp singlet in the presence of the tweezer, thus indicating fast exchange of the tweezer between binding sites at room temperature on the NMR timescale.

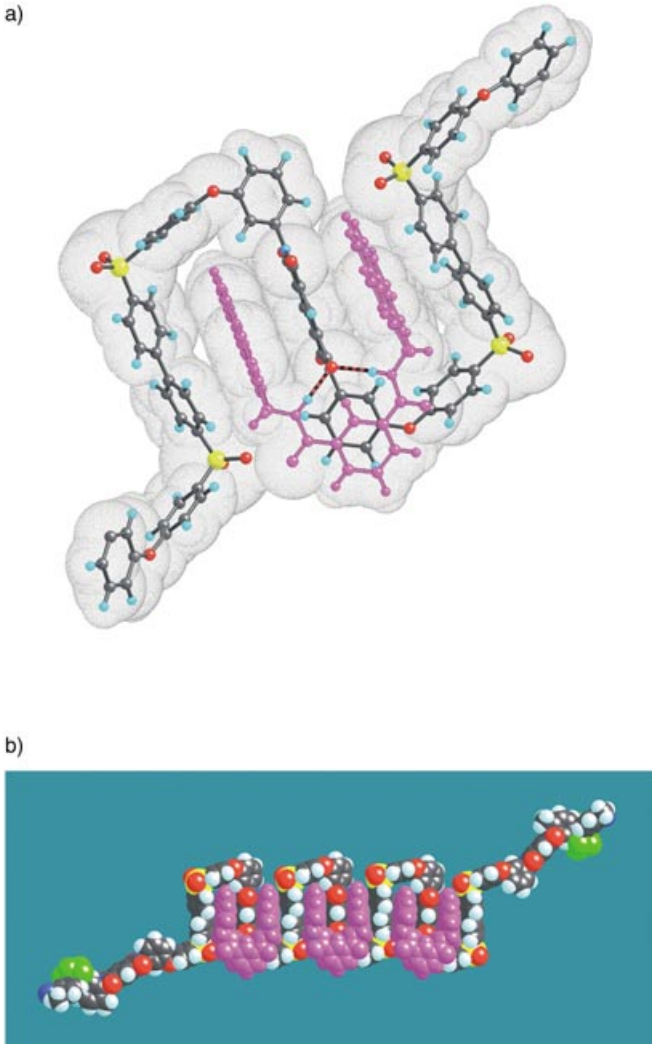
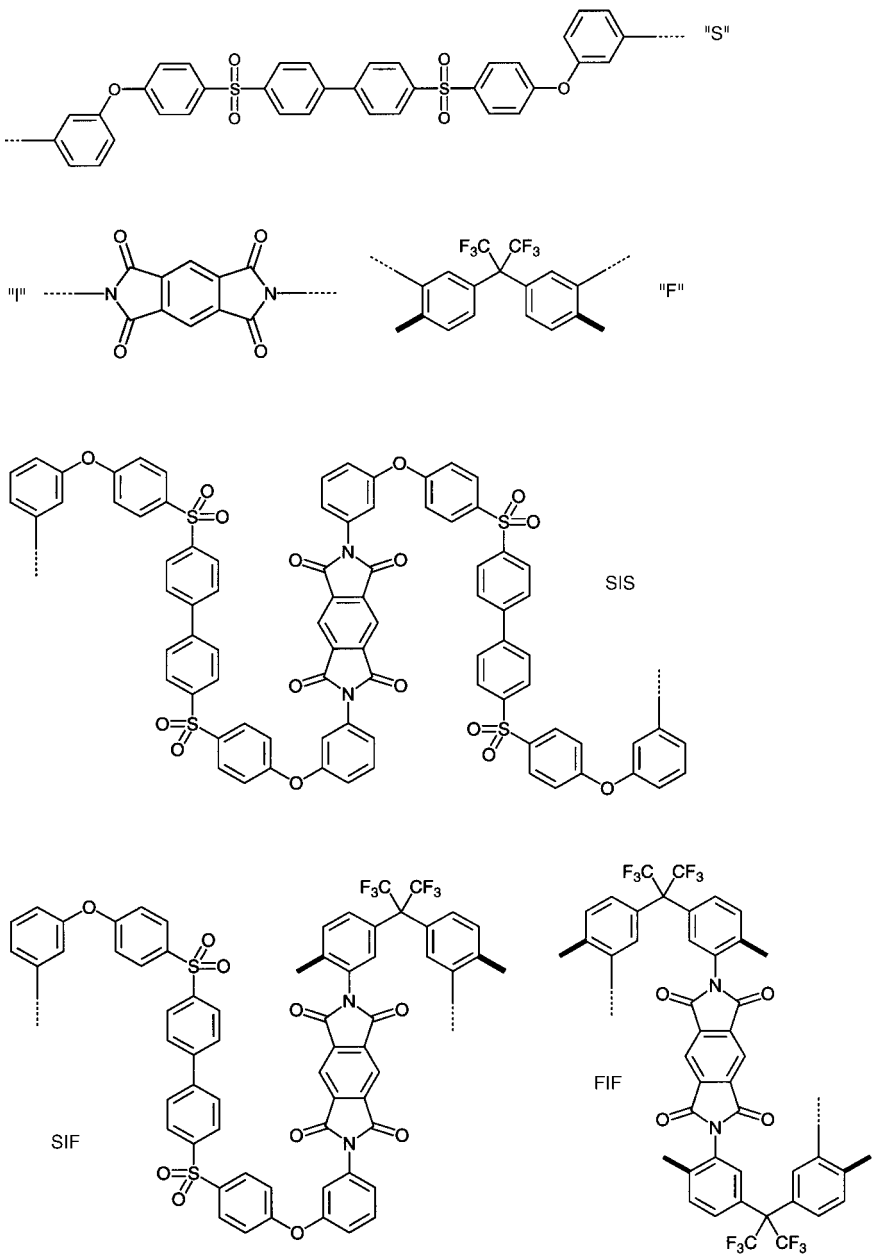


Figure 1. a) Energy-minimized model (re-parametrized Dreiding-II force field) for complexation of the molecular tweezer **1** with the polyimide chain-sequence SIS (see Scheme 1). Tweezer molecule **1** is shown in magenta; van der Waals surfaces are shown in gray; hydrogen bonds between an imide carbonyl oxygen atom and tweezer amide NH groups are shown in red and black. b) Energy-minimized space-filling model of a copolyimide chain showing triply adjacent binding of tweezer **1** (magenta) to the polymer chain-sequence FI-SISISIS-IF.

The monomer sequence that binds a single tweezer molecule contains 13 aromatic rings (Figure 1 a and Scheme 1), but can be simply represented by the string SIS (I is the pyromellitimide binding site and S are the flanking diamine residues, each consisting of six aromatic rings, which upon chain-folding provide additional binding to the tweezer). If it were possible to introduce a diamine comonomer that inhibited, rather than promoted, π-stacking between the imide and the tweezer, then the polymer-sequence recognition and the sequence-specific binding might well be achieved. Computational modeling studies suggested that even a single *ortho* methyl group adjacent to the pyromellitimide unit could lead to significant steric inhibition of π-stacking between the polymer and the tweezer. For this reason, statistical copolymers **7** ($M_n \approx 60 \times 10^3$ Da) and **8** ($M_n \approx 100 \times 10^3$ Da) were synthesized by co-condensation of pyromellitic dianhydride with diamine **4** and with the *ortho*-methylated diamine **6** (chosen for its ability to maintain polymer solubility), in 70:30 and 50:50 mole ratios, respectively.[17] Pyromellitimide resonances in the ^{1}H NMR spectra of copolymers **7** and **8** (Figure 2) are virtually superimposable,[18] thus indicating that in the absence of external interactions the chemical shifts of these resonances are largely insensitive to sequence effects.

If the fluorinated comonomer **6** is defined as F, the possible imide-centered three-component sequences present in these copolymers can be represented by the strings FIF, SIF, FIS, and SIS (Scheme 1). As noted above, modeling



Scheme 1. Imide-centered sequences in a statistical copolyimide based on diamines **4** and **6**. Methyl substituents are shown in bold; in the sequence FIF they can adopt non-interconverting *syn* and *anti* conformations relative to the imide unit. The sequence FIS (not shown) is degenerate with SIF in the present context.

studies indicated that imide groups in the sequences FIF, SIF, and FIS, where one or both aromatic rings adjacent to the imide carry *ortho* methyl groups, should bind only rather weakly to tweezer molecule **1**. However, the string SIS, as shown for homopolymer **3**, binds very strongly to **1** so that, even in a copolymer, the tweezer should induce a marked upfield shift of the imide resonance arising from this sequence. If sequence-recognition and specific binding did indeed occur in a copolyimide of diamines F and S, the imide resonance should separate into two peaks, one at higher field representing the strongly bound sequence SIS and one at lower field representing the much more weakly bound sequences FIF, SIF, and FIS (the latter two sequences being degenerate in this context).

As shown in Figure 2, these predictions were confirmed by experimental data. To avoid possible intermolecular effects between polymer chains, ^{1}H NMR spectra were run at low polymer concentrations, typically 4 mM with respect to the pyromellitimide unit. Addition of tweezer **1** to solutions of copolymers **7** and **8** led to the imide resonance separating into "shifted" and "unshifted" peaks, with integrated intensities of about 1:1 and 3:1 for **7** and **8,** respectively (Figure 2, see column headed "1:80, tweezer/imide"). This peak separation is observed even at extremely low tweezer concentrations ($<10^{-2}$ moles per mole of imide), where tweezer resonances themselves are scarcely detectable in the spectrum. The sequences FIF, SIF, FIS, and SIS should occur in proportions 0.09:0.21:0.21:0.49 for the 70:30 (S/F) copolymer **7**, and in proportions of 0.25:0.25:0.25:0.25 for the 50:50 copolymer **8**. Relative intensities for "unshifted" (FIF, SIF, and FIS) and "shifted" (SIS) imide resonances are thus predicted to be 0.51:0.49 (1.04:1) and 0.75:0.25 (3:1) for copolymers **7** and **8** respectively, which is in good agreement with the NMR data (Figure 2).

When the molar ratio of the tweezer to imide is increased, the SIS imide resonance in these copolymers moves progressively upfield ($\Delta\delta \approx 2.6$ ppm at a ratio of 4:1) and, astonishingly, separates into three clearly defined peaks (Figure 2, see column headed "4:1, tweezer/imide"). The only reasonable explanation for this is that the tweezer is detecting higher-order sequence information, since there are three distinguishable seven-component sequences of monomer residues which incorporate the key central string SIS. These are FISISIF, SISISIF, FISISIS, and SISISIS, with the second and third sequences being degenerate. Statistically, the relative proportions of these different SIS-centered sequences are 1:4.7:5.4 for random copolymer **7** and 1:2:1 for copolymer **8**. The relative intensities of SIS resonances shown in Figure 2 are fully consistent with these ratios, and comparisons of the spectra for **7** and **8** allow individual signals to be assigned to specific sequences on the basis of their positions and relative intensities. The resonance at highest field thus represents the

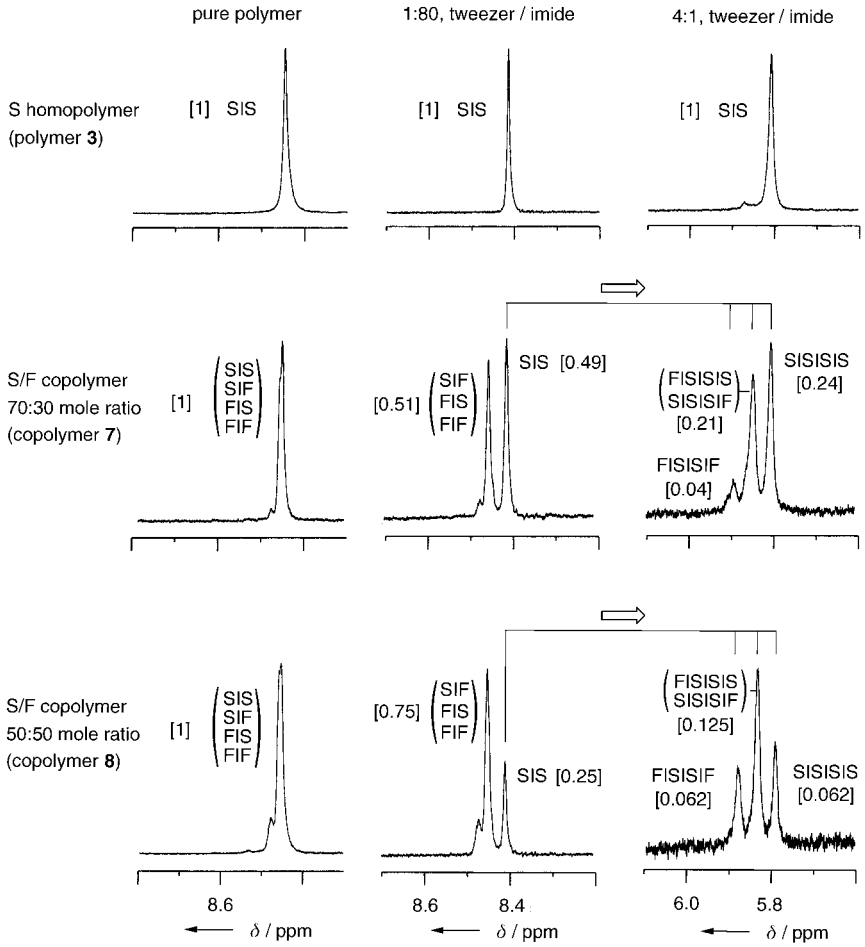


Figure 2. Imide resonances in the ^{1}H NMR spectra of homopolymer **3** and of copolymers **7** and **8** showing the response of imide-centered monomer sequences to the presence of molecular tweezer **1**. Relative integrals predicted from sequence-distribution probabilities are shown in square brackets. Spectra were recorded at 250 MHz in $CDCl_3$/hexafluoropropan-2-ol (6:1 v/v), at a polymer concentration of 4 mM with respect to total pyromellitimide residues. See Supporting Information for full spectra.

central imide residue of the sequence SISISIS, the next highest-field signal corresponds to SISISIF and FISISIS, and that at lowest field to FISISIF.

It may seem remarkable that the tweezer is able to "read" sequence information over such extended stretches of the polymer chain (the string SISISIS for example represents no fewer than 27 aromatic rings) but it may be significant that FISISIF can bind only a single tweezer molecule, whereas SISISIF can bind two, and SISISIS three adjacent tweezers simultaneously, thus leading to the formation of chain-folded and multiply π-stacked clusters (Figure 1b). Multiple adjacent binding is only likely to be significant at high tweezer-to-imide ratios, where additional splitting of the SIS resonance is indeed observed. At high tweezer/imide ratios, even homopolymer **3** (Figure 2, top right) shows an additional, very weak, imide resonance (integral relative to the main imide peak about 0.02:1), which is tentatively assigned to tweezer-bound polymer end groups.

Short-range sequence information in aromatic copolyimides can thus be recognized through the binding of specific monomer sequences to a sterically and electronically complementary molecular tweezer. More remarkably, long-range sequences may also be identified, through the sensitivity of imide ^{1}H NMR resonances to multiple adjacent binding of tweezer molecule **1**. While the issue of "writing" sequence information into the molecular structures of synthetic macromolecules remains to be addressed, it is perhaps worth noting that, in biology, copolymer sequence information is not in fact "written" at all—new DNA sequences arise simply by random mutation and are only ever copied or transcribed. The present work, however, clearly demonstrates that the "reading" of copolymer sequence information in synthetic macromolecules can be achieved through the binding of randomly generated monomer sequences to small molecules of complementary structure.

Received: April 20, 2004

Keywords: molecular recognition · molecular tweezers · pi interactions · polyimides · sequence determination

[1] M. Nirenberg, *Trends Biochem. Sci.* **2004**, *29*, 46.
[2] C. M. Falcon, K. S. Matthews, *Biochemistry* **2000**, *39*, 11074.
[3] J. M. Bujnicki, *Curr. Protein Pept. Sci.* **2003**, *4*, 327.
[4] T. G. Uil, H. J. Haisma, M. G. Rots, *Nucleic Acids Res.* **2003**, *31*, 6064.
[5] Site-specific, but not sequence-specific, binding of macrocyclic receptors to a number of synthetic polymers during polyrotaxane formation has previously been described.[19–21] Polyrotaxanes involving the threading, chain-folding and π-stacking of various aromatic/aliphatic homopolymers with a bipyridinium-based macrocycle have also been reported.[22,23] Polyrotaxane systems[24,25] clearly have enormous potential for displaying sequence recognition, but this does not seem to have been realized in practice so far.
[6] W. C. Still, *Acc. Chem. Res.* **1996**, *29*, 155.
[7] W. C. Still, C. T. Chen, H. Wagner, *Science* **1998**, *279*, 851.
[8] T. Fessmann, J. T. Kilburn, *Angew. Chem.* **1999**, *111*, 2170; *Angew. Chem. Int. Ed.* **1999**, *38*, 1993.
[9] K. Ryan, L. J. Gershell, W. C. Still, *Tetrahedron* **2000**, *56*, 3309.
[10] R. Arenzio, J. D. Kilburn, *Tetrahedron* **2002**, *58*, 711.
[11] H. Wennemers, M. C. Nold, M. M. Conza, K. J. Kulicke, M. Neuburger, *Chem. Eur. J.* **2003**, *9*, 442.
[12] F. M. Menger, A. V. Eliseev, V. A. Migulin, *J. Org. Chem.* **1995**, *60*, 6666.
[13] *Polyimides* (Eds.: D. Wilson, H. D. Stenzenberger, P. M. Hergenrother), Blackie, London, **1990**.
[14] H. M. Colquhoun, Z. X. Zhu, D. J. Williams, *Org. Lett.* **2003**, *5*, 4353.
[15] Computational modeling studies were carried out using Cerius^2 (version 3.5, Accelrys Inc., San Diego), with the Dreiding-II force field reparametrized for aromatic imide and sulfone-based structures.
[16] M. B. Nielsen, J. O. Jeppesen, J. Lau, C. Lomholt, D. Damgard, J. P. Jacobsen, J. Becher, J. F. Stoddart, *J. Org. Chem.* **2001**, *66*, 3559.

[17] The presence of methyl groups *ortho* to the nitrogen atoms in an aromatic diamine do not affect the reactivity of the amino groups towards imide formation, so that random-sequence copolymers can be obtained by co-condensation of *ortho*-methyl-substituted and unsubstituted diamines with aromatic dianhydrides.[26]
[18] The weak shoulder at lower field arises from an unresolved multiplet assigned to non-interconverting *syn* and *anti* arrangements of *ortho*-methyl substituents in FIF sequences.[27]
[19] C. G. Gong, T. E. Glass, H. W. Gibson, *Macromolecules* **1998**, *31*, 208.
[20] S. Choi, J. W. Lee, Y. H. Ko, K. Kim, *Macromolecules* **2002**, *35*, 3526.
[21] D. Tuncel, J. H. G. Steinke, *Macromolecules* **2004**, *37*, 288.
[22] G. J Owen, P. Hodge, *Chem. Commun.* **1997**, 11.
[23] P. Hodge, P. Monvisade, G. J. Owen, F. Heatley, Y. Pang, *New J. Chem.* **2000**, *24*, 703.
[24] S. A. Nepogodiev, J. F. Stoddart, *Chem. Rev.* **1998**, *98*, 1959.
[25] K. Kim, *Chem. Soc. Rev.* **2002**, *31*, 96.
[26] T. M. Wu, S. Chalun, J. Blackwell, S. Z. D. Cheng, Z. Q. Wu, F. W. Harris, *Polymer* **1995**, *36*, 2123.
[27] K. D. Shimizu, T. M. Dewey, J. Rebek, *J. Am. Chem. Soc.* **1994**, *116*, 5145.

