

DNA Nanotechnology Technique for Monodisperse Synthesis

Galen T. Pickett

College of Natural Sciences and Mathematics, California State University Long Beach,
1250 Bellflower Blvd., Long Beach, California 90840

Received June 22, 2006; Revised Manuscript Received October 6, 2006

ABSTRACT: I consider the synthesis of high molecular weight, monodisperse polymers starting from small, monodisperse chains end-labeled with DNA ligands. When these ligands are appropriately designed, a two-stage mixing protocol results in a solution of chains of twice the original molecular weight. The process can be repeated to produce larger and larger chains, of essentially biologically precise molecular weights. The design of precision branched, di-, tri-, and multiblock copolymers is possible under a slight generalization of protocol.

1. Introduction

Today, wonderful control is being exercised over the synthesis of complex polymer molecules in anything from microscopic up to bulk quantity.¹ Orderings of individual monomers on these chains ranging from homopolymer, alternating copolymer, multiblock, multicomponent, branched, grafted, cross-linked, are all achievable. The investment in ordering the sequences and architectures of single polymers allows the supramolecular self-assembly of environment-sensitive patterns.^{2,3} Generally, the more complex the pattern and the more complex the needed behavior of the pattern (for instance, how and when will a particle expand or contract), the more complex the needed constituent polymers are, with consequently more stringent tolerances on errors in synthesis. For example, absolute monodispersity is rarely a requirement when designing a material for a desired bulk application, but monodispersity and low error rates are critical to the proper function of naturally occurring enzymes.

When the polymers involved are intended for use in high-precision measurements on model systems, well-characterized polymers with as closely defined properties as possible are required. Here, I put forward a general method for achieving essentially biologically accurate model materials of various geometries, architectures, and compositions.

The essential strategy is to take advantage of the specific DNA base-pair reaction, where single strands of DNA combine to make the familiar double-helix structure, to create a multitude of specific interaction “sticker” sites⁴ that can be synthesized with relative ease and end-attached to flexible polymers at modest prices. Taking advantage of the self-assembly of complementary single-strand DNA ligands is generally referred to as the “DNA nanotechnology” technique and has been used to construct self-assembled DNA cubes,⁵ truncated octahedra,⁶ two-dimensional crystals,⁷ designed patterns,⁸ scaffolds for computing devices,⁹ walking DNA machines,¹⁰ and “dendrimer-like” buckyball encapsulations.¹¹ The application I envision here is far simpler than that conceived in those works but uses this DNA technique in polymer synthesis. The design method for these structures resembles the long sought-after “tinker toy” approach to complex molecule synthesis but in reality has much more in common with so-called “modular origami”^{12,13} where relatively simple folded structures are designed with clever pocket and tab systems to allow a skillful paper folder to assemble geometric sculptures of incredible subtlety.¹⁴ Here, the units are “folded”—that is, synthesized—beforehand and then

self-assembled rather than assembled by hand.¹⁵ The first recipe I show below aims at creating samples of biologically monodisperse linear polymers. The basic two-stage synthesis protocol I describe has the aim of doubling the chain molecular weight at each stage. Then, a relatively simple variation on the protocol is described, showing how to create specific copolymers with complex block arrangements and branches. A particularly important example of such a self-assembled, regularly branched molecule is the flexible-chain analogue of the dendrimer molecule.¹⁶ A significantly open question for small-molecule dendrimers concerns whether they have voids or solidly filled cores at their centers.^{17,18} The original prediction of the open-core dendrimer was based upon a polymer physicist’s toy model in which monodisperse chains are joined regularly at 3-fold junctions. While there are good reasons to believe that this model in fact requires the cores to be filled,¹⁹ experiments on these molecules are lacking. The method stated below could answer definitely a long-standing controversy.

Additionally, taking advantage of the relatively stiff structure of double-stranded DNA, this method will allow the creation of a class of “freely jointed chain” polymers, essentially microscopic rigid rods linked with free joints. These freely jointed chains are predicted to have unusual liquid crystal properties²⁰ and perhaps phases.^{21,22}

After all of these possibilities are discussed, I will offer a few conclusions as well as directions for further work.

2. Monodisperse Homopolymers

Figure 1A shows schematically the basic two-cycle synthesis route I am proposing. The essential strategy is to use the specific DNA base-pair reaction to drive short, well-characterized chain segments to assemble into pairs, and then these pairwise joined chains will be joined together to make 4-fold chains, and so on.

I will begin with a simple protocol that highlights the main idea but suffers in that reaction fragments from previous synthesis cycles can contaminate the end-product, giving a polydisperse sample. I will then discuss a refinement of the protocol that restricts the molecular weight of the final system to be composed of chains whose molecular weights are distributed in powers of two.

Let us begin with small-molecular-weight homopolymers, of a well-defined molecular weight, N , which have been end decorated with two distinct (although short) ligands of single-stranded DNA, a and b as in Figure 1. The specific sequences

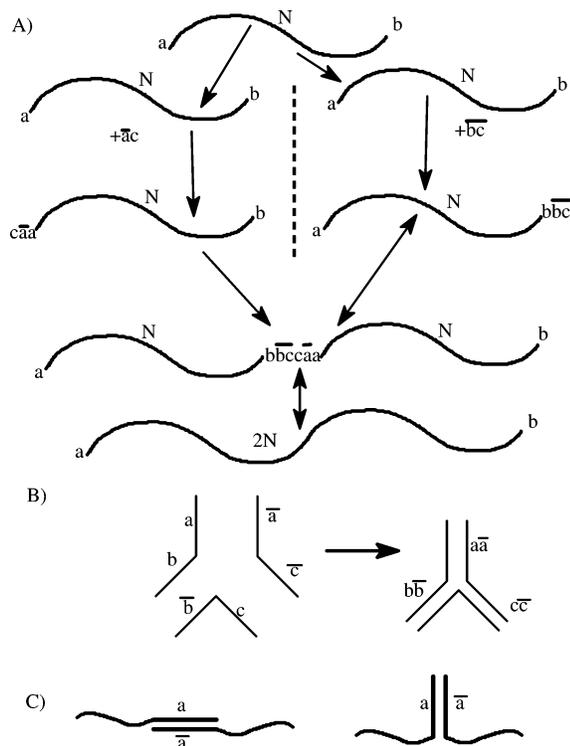


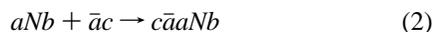
Figure 1. Schematic of the synthesis protocol. (A) Here, monodisperse chains with N monomers are end-labeled with DNA ligands a and b . The system is separated into separate containers and allowed to react with DNA ligands that will *redecorate* half of the chain-ends with ligand c and \bar{c} . When these end-labeled chains are then brought together, the c - \bar{c} complexation reaction will drive the joining of chains to give homopolymers of effectively *twice* the original molecular weight. (B) The trifold junction, tri_{abc} . (C) Main chain (left) and side chain liquid crystal polymers. The main chain polymer is a physical freely jointed chain.

to be chosen for a and for b are arbitrary, but the a sequence of bases should be chosen to not bind well to the b sequence of bases. That is, a and b should not only *not be* complementary base-pair sequences, they should be chosen so that partial base-pairing of a and b should cost free energy. The a sequence has a base-pair complement, \bar{a} , likewise the b sequence complement has a structure \bar{b} , so that the spontaneous reaction



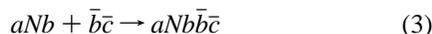
goes to completion.

The system of a , b -end-labeled chains is then divided up into two different containers. To the first container, enough single-stranded DNA with the structure $\bar{a}c$ is added to ensure that the reaction



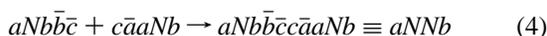
goes to completion. The specific molecular weight N and the energy of the $\bar{a}a$ base-pair reaction can easily be chosen to drive the equilibrium in eq 2 to as small a number as desired of unreacted polymer.

Likewise, the ligand $\bar{b}c$ can be added to drive



to completion.

When the contents of the two containers are thoroughly mixed together, the base-pair condensation of the c and \bar{c} ligands will produce



Chain Distributions

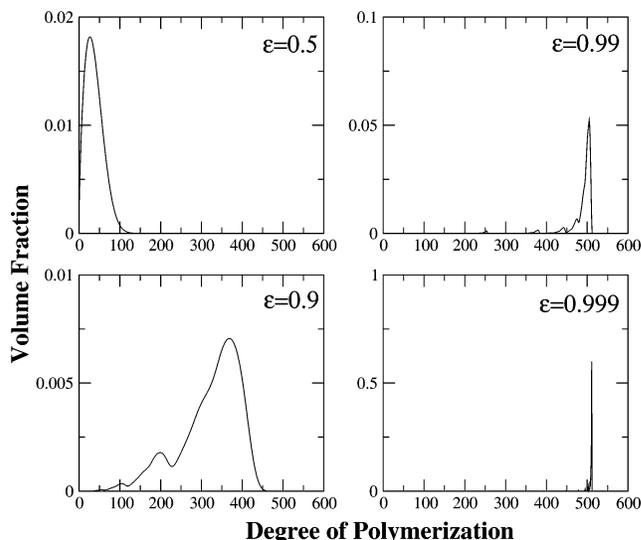


Figure 2. Distributions of chain lengths. In four panels, for $\epsilon = 0.5$, 0.9, 0.99, and 0.99, the distribution of degrees of polymerization of the a - b - c two-stage synthesis cycle is presented after 9 cycles have been performed. As $\epsilon \rightarrow 1$, that is, each cycle reaction is allowed to come more toward completion, the resulting chains are more and more monodisperse, with an enhancement in overall molecular weight approaching a 512-fold increase.

Here, in the final product of the reaction, I have suppressed the internal structure of the joint between the N chains to emphasize the fact that two and only two of the homopolymer chains N are in the final product.

The next step of the synthesis is to take the $aNNb$ chains, divide them into two containers, add the required $\bar{a}c$ and $\bar{b}c$ ligands, and finally mix the results together to produce chains with a structure



That is, each time the cycle is run, there are one-half as many chains, each of which has doubled their molecular weight, a process that can be continued to a desired end-state monodisperse molecular weight.

The above protocol will indeed join chains from the first container to chains of the second container, but the consequences of an incomplete reaction would be to introduce, say, chain fragments with the structure aNb . In fact, the master equation for this two-stage chain growing process is easy enough to write down:

$$f_n^0 = \delta_{n,1} \quad (6)$$

$$f_n^{i+1} = (1 - \epsilon)f_n^i + \epsilon \sum_j f_j^i f_{n-j}^i \quad (7)$$

where ϵ is a measure of the efficiency of the reaction (which is assumed to be the same for each cycle of the reaction), $\delta_{n,1}$ is the Kronecker delta function, and f_n^i is the volume fraction in the i th iteration of the species consisting of n of the homopolymer units. As in Figure 2, when each reaction can be counted on to go to 99% completion, the distribution of molecular weights in the finished sample is weighted heavily toward the target molecular weight, but there is a long trail of lower molecular weight fragments contaminating the system. An interesting point, however, is that molecular weight distribution is peaked at the largest possible molecular weight under this scheme.

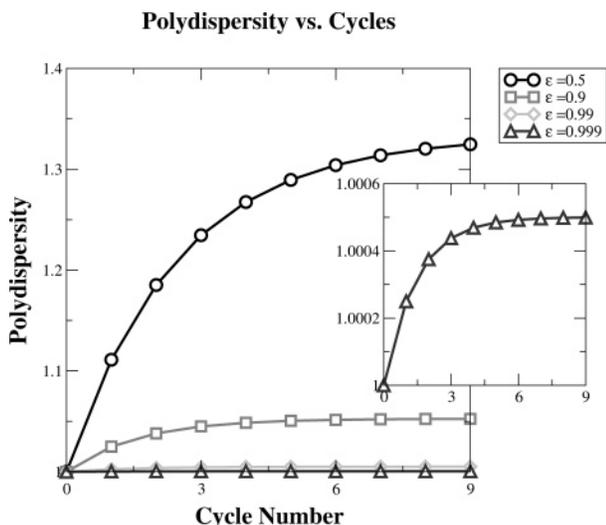


Figure 3. Polydispersity. The polydispersity (M_w/M_N) at mixing cycles up through 9 for various ϵ are shown. Inset is the polydispersity of $\epsilon = 0.999$.

Here the efficiency of the reaction, ϵ , determines the overall molecular weight distribution at the end of the synthesis. With $\epsilon = 0.5$, there are always plenty of unreacted chain fragments in the solution, and a typical synthetic polymer molecular weight distribution results. When each stage of the reaction is required to go to 99.9% completion, the distribution is essentially monodisperse, with approximately one-half of the total volume fraction of chains in the system achieving the target molecular weight.

The polydispersity, the ratio of the weight-average to number-average molecular weights

$$M_w/M_N \equiv \frac{\sum f_n^i n^2}{(\sum f_n^i n)^2} \quad (8)$$

produced in the synthesis scheme is presented in Figure 3. Even with $\epsilon = 0.5$, M_w/M_N approaches a respectable value of 1.3, whereas typical high molecular weight synthesis can easily result in polydispersities in the range 2–3. As $\epsilon \rightarrow 1$, the polydispersity saturates at values closer and closer to 1, finally achieving an impressive value of 1.0005 for $\epsilon = 0.999$. Essentially biological precision has been achieved.

In Figure 4, I show the polydispersity after nine cycles of synthesis have been completed at various values of ϵ . For $\epsilon \rightarrow 0$, the system maintains very nearly $M_w/M_N \approx 1$, but this is a result of the fact that very few chains have reacted at all. With limited reactivity, most chains remain in the reservoir of unreacted “monomer”, and chains grow by a Poisson process, randomly adding chains as time goes on. As we restrict the reaction to more and more follow our two stages of segregation and combination, we get a transition at $\epsilon \approx 0.2$ from a system dominated by Poisson growing chains to chains that are growing by doubling. As $\epsilon \rightarrow 1$, the lower molecular weight chains are suppressed entirely in favor of doubling chains, and we attain a monodisperse, high-molecular weight system where the overall chain length is $512N$.

It is important to note, however, that this monodispersity can only be achieved if we are willing to allow each reaction to proceed to a specific level of completion, ϵ . The reaction times to achieve specific ϵ will increase dramatically as the cycle number increases. As the molecular weight of the chains increases, the self-diffusion time will also increase. Even if the

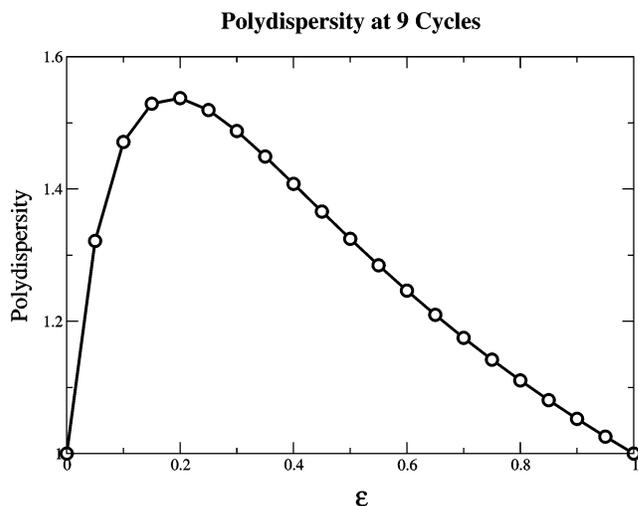


Figure 4. Polydispersity at 9 cycles. The polydispersity M_w/M_N after 9 synthesis cycles is presented at various ϵ . For small values of ϵ , the system is nearly monodisperse because relatively few chains have been joined together, and the overall average molecular weight of the chains is nearly 1. As $\epsilon \rightarrow 1$, however, the chains grow larger and larger, and all nearly achieve, $M_w = 2^9 N$.

chains never surpass their entanglement threshold, this diffusion time obeys

$$T_i \sim (2^i N)^2 \quad (9)$$

However, as the chains are doubling in molecular weight, a rough estimate for the overall chain growth is

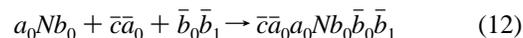
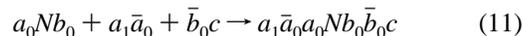
$$\frac{dM}{dt} = \frac{2M}{M^2} \quad (10)$$

where M is the current molecular weight of the chains, leading to an overall synthesis time that is quadratic in the target molecular weight.

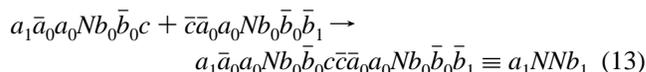
When wait times to achieve a particular ϵ become experimentally untenable, additional filtration and separation steps may be added to the protocol. The basic idea is that the highest molecular weights are going to be easy to separate from the smaller molecular weights in solution.

One way around this difficulty is to complicate the synthesis cycle a bit, so that at each stage of the synthesis there are unique end ligands \bar{a}_n and \bar{b}_n . In this way, the unreacted fragments will drop out of the propagation of chains but will leave a telltale molecular weight doubling signature in any separation process useful for precision calibration of molecular weights.

Again, we start with a system of short, monodisperse polymer end-labeled with DNA ligands a_0 and b_0 . We separate the system into identical subsystems, but this time we add two ligand species to each cycle step:



so that the final cycle reaction becomes



In this case, the lower molecular weight unreacted $a_0 N b_0$ chains have ligands that will be incapable of growing during the next cycle because chain growth in the next cycle will only occur

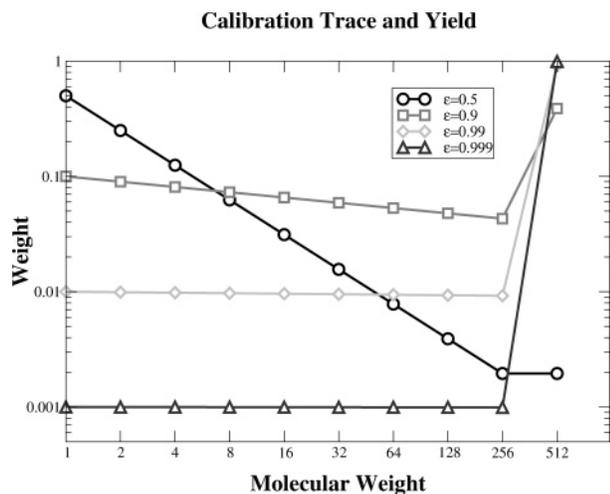


Figure 5. Second protocol chain distributions. In the second protocol, small molecular weight fragments are produced, but only at molecular weights that are power-of-2 multiples of the basic chain molecular weight N . Lines are drawn here as a guide to the eye only. As $\epsilon \rightarrow 1$, most of the weight in the system is retained in a monodisperse peak at $M_N = 512$, but the trailing fragments are present and could serve as a precise molecular weight calibration tool.

for chains with the correct a_1 and b_1 ligands. The only requisite now is that the end-decoration reactions, eqs 11 and 12, must run to near completion. The rate-determining step will still be the joining of large chains by end-association, as in the original two-stage protocol above. Thus, there is no particular speedup of the chain-growing reaction in the second protocol. The advantage in the second protocol is that even if that chain-joining reaction is characterized by some ϵ as above, the growing chains at each stage will all have the same molecular weight.

If, as above, we use ϵ to stand for the degree of completion of the cycle reaction, then we can again ask about the overall polydispersity of the sample of chains. In the second protocol, after i cycles have been performed, only chains with molecular weights which are powers of 2 are present. That is f_n^i vanishes unless $n = 2^m$ for $1 < m < i$. In fact

$$f_{2^m}^i = (1 - \epsilon)\epsilon^m \quad \text{for } 1 < m < i - 1 \quad (14)$$

$$f_{2^i}^i \equiv f_{\max}^i = \epsilon^i \quad (15)$$

The smaller molecular weight chains are confined to molecular weights that are powers of 2 combinations of the end-decorated homopolymers, and the ultimate yield of the highest molecular weight chains takes a very simple exponential form: $f_{\max}^i = \epsilon^i$. In the second protocol, the small-molecular-weight chains produced by incomplete reactions are located conveniently in molecular weight to serve as an extremely precise calibration of, for instance, a gel chromatography column. In contrast, the unreacted chains in the first protocol give rise to a background spectrum of “noise” in the molecular weight distribution. Figure 5 shows this effect in detail for various values of ϵ . For $\epsilon = 0.5$, this protocol produces an exponentially smaller weight of exponentially longer chains, giving a straight line on the logarithmic plot in Figure 5. For $\epsilon \geq 0.9$, however, the chain fragments occur with nearly the same overall concentration, and nearly the entire weight of the sample is concentrated at the highest molecular weight, in this example, $M_N \approx 512N$.

The polydispersity of this second protocol sample can also be readily calculated from eq 15. As shown in Figure 6, the polydispersity can approach quite high values ($M_w/M_N \approx 25$

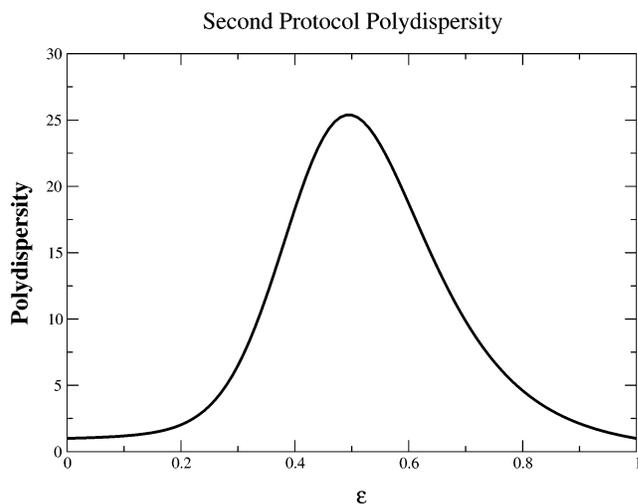


Figure 6. Second protocol polydispersity. The second protocol polydispersity at cycle 9 is presented as a function of ϵ .

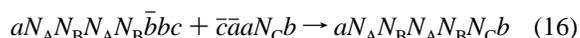
for $\epsilon = 0.5$) but rapidly approaches unity when $\epsilon \rightarrow 1$. The large values of the polydispersity, even for $\epsilon = 0.5$, do not tell the whole story, however. In each case, the molecular weights in the sample only occur at well-separated values of molecular weight, and it is far easier to separate, say, a mixture of $M_N = 256$ and $M_N = 512$ chains than it would be to separate, say, $M_N = 512$ from $M_N = 511$. In this case, even with large values of polydispersity, the system is well characterized as a mixture of pure molecular weights. In this case, too, nearly biological control over the molecular weight of synthetic polymers may be achieved.

3. Multiblock, Multigraft Copolymers

Generalizing either protocol above to creating multiblock copolymers of a specific architecture is straightforward. Diblock copolymers are composed of N_A and N_B monomers of type A and B, respectively, joined end-to-end. The remarkable properties of these copolymers arise from the statistical incompatibility of the A and B blocks, driving a microsegregation in the bulk and various interesting patterns in thin films.

3.1. Block Copolymers. To construct diblock copolymers, all that would be required is several cycles of the synthesis protocol be run to create a - and b -end-labeled homopolymers of molecular weights N_A and N_B , respectively. Given the geometric increase in overall molecular weight, it will be necessary to choose the molecular weight of the spacer N chains with care, but this certainly does not pose a significant difficulty.

By taking advantage of the $c\bar{c}$ base-pair reaction, triblock and general multiblock copolymers of arbitrary and well-controlled architectures can be attained. For instance, if we start the protocol off with separated but labeled N_A and N_B chains, after the first cycle of the protocol, we will have chains with the architecture $aN_A N_B b$. When the protocol is run on this mixture of diblock copolymers, the next stage will produce $aN_A N_B N_A N_B b$ chains. Such alternating block copolymers are predicted to be particularly good interface strengthening agents for polymer blends.^{23,24} That is, an alternating block copolymer will be synthesized. If instead of separating these chains into identical systems and relabeling they are mixed with labeled N_C subchains (made of a third type of C monomer), we will have synthesized:



The range of copolymers that may be synthesized in this technique is startling.

3.2. Multigraft Copolymers. Additionally, this DNA technique will allow the construction of regularly branched polymers, again with remarkable control over architecture, monomer sequencing, and overall chain lengths. The basic architectural feature of regularly branched polymer is an n -fold junction of chains. The 3-fold junction, fortunately, is the basic structure that was invented to construct DNA polyhedra such as the cube and truncated octahedron.^{4–6} Three mutually associating DNA chains are allowed to complexity:



Star-block copolymers,²⁵ dendrimer copolymers,^{26,27} and dendrimer–homopolymer copolymer “tadpoles”²⁸ can all be readily synthesized using the above two-stage synthesis protocol, with the relatively small change of adding new types of DNA ligands in the end-alteration stages.

3.3. Liquid-Crystal Polymers and Freely Jointed Chains. To this point, I have made the implicit assumption that the flexible spacer chains, N , are the primary structural elements of the chain architecture. Thus, the connecting DNA ligands can be considered to be a small portion of the overall chain composition and therefore only a small perturbation on any material property. However, the joint complexes are composed of double-helix DNA, a material that is charged and relatively stiff. If the ligands are made much longer than a few tens of base pairs, the “connecting” segments behave much as rigid rods, and the flexible spacers are more like freely deformable joints between these rigid rods. Whether the chains then behave as main-chain liquid crystal polymers or side-chain liquid crystal polymers is entirely a function of how the a and b ligands are chosen, as in Figure 1C. This method thus has the promise of creating a physical substantiation of the “freely jointed chain”—to this point a very well-studied toy model. Thus, the DNA construction method can produce copolymers with side and/or main chain characteristics.

4. Conclusion

I have proposed what should prove to be a uniquely robust method for the precision-crafting of complex molecules in, if not industrial capacity, then at least in yields sufficient for several precision experiments. The method takes advantage of the remarkably specific DNA base-pairing interaction to guide the assembly of chain segments into progressively larger and larger structures. The synthesis of monodisperse chains, of well-

characterized monodisperse multiblock copolymers, and branched copolymers have been presented. I hope that, given extremely well-characterized samples with complex geometries, several theorists’ models and simplifying assumptions can be put to rigorous experimental tests. Additionally, the theorist’s “freely jointed chain model” is given a physical substantiation.

References and Notes

- (1) Hong, J.; Wang, Q.; Fan, Z. *Macromol. Rapid Commun.* **2006**, *27*, 57.
- (2) Ding, H.; Wu, F.; Huang, Y.; Zhang, Z.-R.; Nie, Y. *Polymer* **2006**, *47*, 1575.
- (3) Schilli, C. M.; Zhang, M. F.; Rizzardo, E.; Thang, S. H.; Chong, Y. K.; Edwards, K.; Karlsson, G.; Muller, A. H. E. *Macromolecules* **2004**, *37*, 7861.
- (4) Winfree, E.; Liu, F. R.; Wenzler, L. A.; Seeman, N. C. *Nature (London)* **1998**, *394*, 539.
- (5) Chen, J.; Seeman, N. C. *Nature (London)* **1991**, *350*, 631. Database: Compendex.
- (6) Zhang, Y. W.; Seeman, N. C. *J. Am. Chem. Soc.* **1994**, *116*, 1661.
- (7) Winfree, E.; Furlong, L.; Wenzler, L. A.; Seeman, N. C. *Nature (London)* **1998**, *394*, 539.
- (8) Rothmund, P. W. K. *Nature (London)* **2006**, *440*, 297.
- (9) Rothmund, P. W. K.; Papadakis, N.; Winfree, E. *PLoS Biol.* **2004**, *2*, e424.
- (10) Sherman, W. B.; Seeman, N. C. *Nano Lett.* **2004**, *4*, 1203.
- (11) Li, Y. G.; Tseng, Y. D.; Kwon, S. Y.; D’Espaux, L.; Bunch, J. S.; Mceuen, P. L.; Luo, D. *Nat. Mater.* **2004**, *3*, 38.
- (12) Fuse, T. *Unit Origami: Multidimensional Transformations*. Japan Publications (USA), 1990 ISBN 0870408526.
- (13) Gurkewitz, R.; Arnstein, B. *Multimodular Origami Polyhedra: Archimedean, Buckyballs and Duality*, Dover: 2003 ISBN 0486423174.
- (14) Particularly see: “Cosmosphere”. Kawamura, M. In Lafosse, M. *Origamido: The Art of Folded Paper*, Rockport Publishers: 2000 ISBN 1564966399.
- (15) The design of three-dimensional patterns as well as scaffolded patterns has been referred to as “DNA-origami”. Thus, the general DNA nanotechnology methods I make use of here are related, at least on a conceptual level, to that more specific “origami” term.
- (16) Tomalia, D. A.; Baker, H.; Dewald, J.; Hall, M.; Martin, S.; Roeck, J.; Ryder, J.; Smith, P. *Polym. J.* **1985**, *17*, 117.
- (17) Lescapec, R. L.; Muthukumar, M. *Macromolecules* **1990**, *23*, 2280.
- (18) de Gennes, P.-G.; Hervet, H. *J. Phys. (Paris)* **1983**, *44*, L351.
- (19) Zook, T. C.; Pickett, G. T. *Phys. Rev. Lett.* **2003**, *90*, 15502.
- (20) Pickett, G. T.; Witten, T. A. *Macromolecules* **1992**, *25*, 4569.
- (21) Pickett, G. T.; Schweizer, K. S. *J. Chem. Phys.* **1999**, *110*, 6597.
- (22) Pickett, G. T.; Schweizer, K. S. *J. Chem. Phys.* **2000**, *112*, 4881.
- (23) Pickett, G. T.; Jasnow, D.; Balazs, A. C. *Phys. Rev. Lett.* **1996**, *77*, 671.
- (24) Pickett, G. T.; Jasnow, D.; Balazs, A. C. *Trends Polym. Sci.* **1997**, *5*, 128.
- (25) Milner, S. T. *Macromolecules* **1994**, *27*, 2333.
- (26) Rios, G. E.; Pickett, G. T. *Macromolecules* **2003**, *36*, 2967.
- (27) Grason, G. M.; Kamien, R. D. *Phys. Rev. E* **2005**, *71*, 051801.
- (28) Pickett, G. T. *Macromolecules* **2002**, *35*, 1896.

MA061405S