Use of Cohesive-Ended DNA "Monomers" as Models for the Study of Polymeric Systems: A Perspective†

Homer Jacobson

Department of Chemistry, Brooklyn College, Brooklyn, New York 11210

Received April 2, 1991; Revised Manuscript Received September 5, 1991

ABSTRACT: Advantages of using cohesive-ended double-strand DNA molecules as monomers in experimental polymer studies are pointed out. These include the ability to simulate a variety of linear and cross-linked polymeric systems, wide range of monomer sizes, accurate and powerful synthetic methods, low excluded volumes, reaction control with temperature regulation, and direct interpretation by EM and electrophoresis. Desirable work in the areas of distributions of catenanes and knots, nonequilibrium systems, and cross-linked systems containing catenanes is called for. Ends are most effective with 10–20 bases, which give convenient melting points for the cohesion reaction. Monomers should be in the 3–5 kilobase pair range, with non-self-complementary ends, to maximize ring and catenane formation without effects of rotational alignment or need for wormlike, rather than free-link, statistics. Shorter monomers allow the study of DNA-specific polymer problems, particularly those involving small rings.

General Considerations

An extremely useful and controllable experimental system for measuring equilibrium or nonequilibrium distributions of chains, rings, catenanes, and knots can be set up in a polymerizing mixture in which the "monomers" are double-strand DNA segments with natural or synthetic cohesive ends (CEs) of reasonable size. As we are aware, CE "bonds" are merely geometric groupings of H bonds and London attractions, hardly qualifying the resulting aggregates as condensation polymers or as double-bond type addition polymers. In distributions, kinetics, and statistics, however, they mimic condensing mixtures, reversible unless the cohesion is made permanent by ligation or cross-linking. This paper will preview possibilities for extension of polymer chemistry using CE DNA.

Natural CE DNAs, like that of λ-phage, and restriction-generated CEs have long been known;[1,2] both have been studied extensively.[3,4] Polymeric studies with CEs were first made by Wang and Davidson,[5,6] who measured the equilibrium and kinetics of the transition between monomeric linear and cyclic forms of λ-phage DNA. Further studies were performed by Baldwin and co-workers[7] whose work on restriction-cut DNA segments of variable size bearing short CEs has supplied important information on ring closure of DNA, verification that DNA observes wormlike chain statistics[8,9] (or with longer chains, its asymptotical result, Gaussian chain statistics), and has given evidence for a periodic rotational alignment factor.[10] Extensive work has been reported on the kinetics of ligation of restriction ends.[11-13] Cosmids, hybrid cycles with λ-CEs and plasmid DNA inserts,[14,15] are also well-known and have been utilized for DNA library packaging. With the explosive growth in DNA synthetic and recombination techniques, however, more detailed studies can now be undertaken, using monomers and CEs with far more variable sizes and compositions. For such studies, λ is too large, restriction fragment CEs too weakly bonded, and cosmids limited by their CEs.

Such polymeric systems can do the following: (1) Use homogeneous monomers with controlled parameters, made with standard recombination-technology methods, suitably modified to make cohesive ends. (2) Attain the desired overall fraction of reaction at equilibrium (p) by temperature regulation. (3) Irreversibly polymerize using ligases, or molecules which cross-link at the cohered CEs. (4) Produce closed circular (doubly) nicked double-stranded (CCND) DNA in high yield, along with equilibrium amounts of catenanes and knots, by annealing at high dilution. (5) Produce distributions of (nicked) chain polymers with low content of CCND DNA, by equilibrating at high concentrations. (6) Produce closed-circular covalent double-stranded (CCCD) DNA after total ligation. (7) Produce negatively supercoiled CCCD DNA mixtures by using gyrase after ligation, and positively supercoiled mixtures by ligation in the presence of intercalators with subsequent removal. (8) Produce cross-linked polymers, using stabilized DNA three-way or four-way junctions with CEs.

An equilibrium experiment starting with any form or mixture of monomers, chain or rings, must be run at a practical annealing temperature. The monomer with its CEs is the primary variable. Varying temperature about the cohesion midpoint (T_m) causes the value of p to run the gamut between unity and zero and thus produces a wide set of molecular distributions. Varying the concentrations similarly gives a range of ring/chain/catenane ratios.[16]

Given these variables, DNA chain flexibility[17] is the primary contributor to size distributions. Its high stiffness furnishes a minimal though significant contribution of chain thickness to steric hindrance.[18] Solvent variability, particularly ionic strength, influences the thermodynamics of cohesive-end pairing[6] and, less dramatically, the effective chain diameter[19] and persistence length;[20] after ligation, effects of supercoiling[21] become apparent.

Cohesive ends in the 10–20-bp range (bp = base pair) (12-bp λ-ends melt in the 50 °C range at low salt concentration[5]) seem the most practical size, avoiding T_m values that are too high or too low. Low T_m causes low p at room temperature, and high T_m threatens stability of the main helix at reasonable values of p.

The ideal monomer length will vary with experimental requirements. Both theoretical and experimental ring-closure optima are at about 500 bp.[7,9,22] At 3 kbp (about 10 Kuhn segments[17]), it is within 1 order of magnitude of optimum in producing DNA rings, large enough for the rotational alignment factor to become negligible in ring

† Editor's note: This unusual paper contains neither new experimental results nor theory, but the editors and the reviewers believe that it will stimulate the imagination and broaden the perspective of its readers.

0024-9297/91/2224-6712$02.50/0

closure[10] and for simple Gaussian chain statistics to become nearly equivalent to those of wormlike chains.[20,22] While the independence of cohesion–decohesion enthalpies and entropies on molecular size is not absolute, we can expect constancy when strains or steric factors are not substantial. For the cited ring size, neither will be large.

Production of monomers is possible over an extreme range of sizes, downward to a few tens of base pairs, by at least up to the 50-kbp characteristic of λ and the like. Many methods for producing these are apparent to workers with recombinant and synthetic DNA techniques. They include, among others,[23] extension of restriction-cut DNAs by sealing matched synthetic oligonucleotide adaptors with ligase[24,25] or by pairing pieces cut from two batches of CCND DNA at two different loci.[26] With "palindromic" (self-complementary) ends, the monomers follow case I[27] (polyglycol type) statistics; otherwise, polarized case II (polypeptide type) polymerization results. Biological DNA CEs give rise to case II monomers; but the mostly palindromic restriction-cut CEs produce case I monomers. Man-made types can be produced[23] for either category and could create case III (nylon copolymer type) mixtures. Chain and ring distributions for these types have long been familiar. For statistical reasons, case II monomers lead to 2 times as many rings[27] and 4 times as many catenanes[28] as do case I. For the study of topologically complex conditions, cases I and II provide fewer theoretical problems. Similar choices of reaction types apply to combinations including cross-linked polymers.

Equilibrium Distributions

Results of the cited work on CE polymerization indicated small quantities of catenanes as probable products and a range of utility for classical ring–chain equilibrium theory.[27] Yet, for more than 2 decades, measurements of chain/ring/catenane distribution have not been made with efficient ring and catenane producers nor have they been done over a range of monomer or polymer sizes.

Chains are primarily in Flory distributions determined by p; cyclic polymers, catenanes, and knots are determined by concentrations and entropic factors. Distributions of molecular species are measurable with gel electrophoresis[29] using electron microscopy to index ladders or two-dimensional patterns and to identify complex species.[30,31] These can furnish the true steric factors of catenane and knot formations, which are, in the absence of firm analytical models, still best settled by observing catenation[32] and knotting.[30,33,34] Recent but optimistic theory[16,28] suggests that there are catenanes to be found in reasonable quantity; other theories[35-37] give rough quantitative estimates for catenanes and knots.

Species with three or more CEs are needed to cross-link the linear monomers described. Stable three- and four-arm DNA junctions have been described and characterized by Kallenbach, Seeman, and co-workers.[38,39] Four three-armed junctions, provided with CEs, were ligated into a cyclic tetramer bearing four arms.[40] While in this case the CEs differed, such junctions can be constructed with identical palindromic ends and made part of a mix of tri- and tetrafunctional monomers. Equilibrium systems have been observed at various cross-linking ratios and reaction fractions, as treated by the classic theories of Flory and Stockmayer and by more recent treatments describing both equilibrium and irreversible polymerizations.[41-43]

Under conditions which would be expected to produce a substantial catenane fraction, we do not predict[16] a catenane "chain mail" gel from cyclic polymers in equilibrium with linear chains at any attainable value of p.

The relative roles of cross-linking and interpenetration of rings may be observable in combined systems. Interpenetration, i.e., topological cross-linking, is a feature of catenation. Effects of catenative "trapping" and interpenetrating networks have been investigated with cyclic poly(dimethylsiloxane).[44,45] They should be even more advantageously observable with cross-linked and cyclic systems from stiff DNA-based monomers. Catenation accelerates the gel point caused by cross-linking, while noninterpenetrating rings delay it.

Quenching and Irreversible Distributions

Distributions at equilibrium must have both cohesion and decohesion reactions stopped before measurement. A temperature quench is the simplest and most desirable; the reaction of λ ends effectively stops[5,6] at 4 °C. Thus, CEs at least as long as λ's should cease both cohesion and the slower decohesion at low temperatures.

Comparison of distributions of nearly identical polymer systems, one in equilibrium and one irreversibly polymerized, is possible if reverse reaction is stopped during cohesion, e.g., by using fast ligation or cross-linking. The relative dependence of cyclization and catenation on the reversibility of the reaction can be explored. Recent interest in distributions of rings and chains resulting from nonequilibrium reactions, especially irreversible ones, has led to numerous experimental and theoretical treatments of this problem.[46-48] With CEs, one can make direct experimental comparison with equivalent monomers.

In the irreversible case, time rather than temperature determines the final distribution. For best comparison, they should both be studied at similar temperatures, as well as at the same extent of reaction.

Ligation is usually slower than the constant breakup and reuniting of short (i.e., restriction) ends. Cited work with short ends[7,10] used ligation perturbing such a reversible steady state. For λ-DNA at ionic strength 0.13 M, the Wang–Davidson data (eq 18 in ref 5 and their free energy estimates) leads to a first-order breakup rate k_b given by $\log k_b\ (\mathrm{min}^{-1}) = -2.489 \times 10^4/T + 74.8$.

This gives a time constant of 1 min at 60 °C (about 9 °C above the T_m) and time constants of 4×10^5 and 10^8 min at 37 and 27 °C. As ligations are generally conducted with time constants of 10–10^3 min, conditions avoiding significant breakup of cohered but unligated ends during reaction can be envisioned easily.

Optimizing the ligation/decohesion ratio thus mandates both sufficiently long CEs and optimal temperature and ionic strength for ligation. For T4 ligase, this means about 30 °C and 0.05 mM, at which point cohesion for λ-ends is quite slow and decohesion negligible. Lengthening the ends will slow the forward rate unacceptably, and shortening them will risk a too rapid decohesion; the optimal end lengths can be established with pilot studies. *Escherichia coli* ligase, being reversible in the presence of AMP, should be avoided. Control studies using a ligase system missing a cofactor can show whether the presence of inactive ligase affects the cohesion equilibrium.

Using a higher temperature for the equilibrium reactions than the irreversible ones might prove necessary; at the same values of p, comparison of distributions is still possible. Where decohesion is negligible, ligase is unnecessary and nonequilibrium distributions can be compared with those at equilibrium with the same p at higher temperatures. The most interesting reaction regions, however, are those at high extents of reaction and ring formation, i.e., where decohesion is most significant. Here, theoretical

results are most difficult to derive, and the distributions with the two reaction types are expected to diverge.[46]

DNA-Specific Problems

The suggestions above deal with general polymer problems. Many further experiments are specific to either DNA, cohesive ends, or tiny DNA rings. Measurements made with short cyclizing monomers using cohesive ends longer than restriction-cut ones will improve the detectability of the smallest and most strained DNA rings and the specification of chain statistics and moduli. Below ca. 200 bp, chain stiffness fights cyclization[7] but smaller rings have been created by heavily kinking the chain with special proteins.[49,62]

DNA has sequence-dependent proclivities for isotropic or anisotropic curving. "Fishhook" species[50] show some of these curving sequences to be natural; other natural DNAs show double-strand bodies interrupted by single-strand segments.[51] Determination of chain parameters and statistics for chains with deletion kinks[52] or permanent bends[53] has been of interest. These are constructible from CE monomers containing specific bender or kinker sequences or single-strand or unmatched segments or other synthetic but flexible links on one or both ends. The cohesive-end system is especially suited for experimental observations of DNA made from monomers containing controlled irregularities, the polymers bearing repetitions of these. Current statistical theories have treated the origin of sequence-determined bending[54-58] in detail, but further study of regularly or randomly altered chain DNA should be welcome, as such bends occur in the presence of proteins,[59] drugs,[60] and other chemicals.[61]

Acknowledgment. I gratefully acknowledge support for this project from Grant No. 665128 of the PSC-CUNY Award Program of CUNY and fruitful discussions with Norman Davidson and Bruno Zimm.

References and Notes

(1) Hershey, A. D.; Burgi, E. *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *53*, 325.
(2) Mertz, J. E.; Davis, R. W. *Proc. Natl. Acad. Sci. U.S.A.* **1972**, *69*, 3370.
(3) Hendrix, R., et al., Eds. *Lambda II*; Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, 1983.
(4) Roberts, R. J. *Nucleic Acids Res. Suppl.* **1990**, *18*, 2331.
(5) Wang, J. C.; Davidson, N. *J. Mol. Biol.* **1966**, *19*, 469.
(6) Wang, J. C.; Davidson, N. *J. Mol. Biol.* **1966**, *15*, 111.
(7) Shore, D.; Langowski, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1981**, *78*, 4833.
(8) Gobush, W.; Yamakawa, H.; Stockmayer, W. H.; Magee, W. S. *J. Chem. Phys.* **1972**, *57*, 2839.
(9) Levene, S. D.; Crothers, D. M. *J. Mol. Biol.* **1986**, *189*, 61.
(10) Shore, D; Baldwin, R. L. *J. Mol. Biol.* **1983**, *170*, 957.
(11) Dugaiczyk, A.; Boyer, H. W.; Goodman, H. M. *J. Mol. Biol.* **1975**, *96*, 171.
(12) Legerski, R. J.; Robberson, D. L. *J. Mol. Biol.* **1985**, *181*, 297.
(13) Revie, D.; Smith, D. W.; Yee, T. W. *Nucleic Acids Res.* **1988**, *16*, 10301.
(14) Hohn, B.; Collins, J. *Gene* **1980**, *11*, 291.
(15) Ishiura, M.; Ohashi, H.; Uchida, T.; Okada, Y. *Anal. Biochem.* **1989**, *176*, 117.
(16) Jacobson, H. *Macromolecules* **1988**, *21*, 2842.
(17) Hagerman, P. J. *Ann. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 265.
(18) Klenin, K. V.; Vologodskii, A. V.; Anshelevich, V. V.; Dykhne, A. M.; Frank-Kamenetskii, M. D. *J. Biomol. Struct. Dynam.* **1988**, *5*, 1173.
(19) Hagerman, P. J. *Biopolymers* **1981**, *20*. 1503.
(20) Hagerman, P. J. *Biopolymers* **1985**, *24*, 1881.
(21) Klenin, K. V.; Vologodskii, A. V.; Anshelevich, V. V.; Klishko, V. Y.; Dykhne, A. M.; Frank-Kamenetskii, M. D. *J. Biomol. Struct. Dynam.* **1989**, *6*, 707.
(22) Shimada, J.: Yamakawa, H. *Macromolecules* **1984**, *17*, 689.
(23) Jacobson, H. Unpublished notes.
(24) Haymerle, H.; Herz, J.; Bressan, G. M.; Frank, R.; Stanley, K. *Nucleic Acids Res.* **1986**, *14*, 8615.
(25) Taylor, W. H.; Hagerman, P. J. *Gene* **1987**, *53*, 139.
(26) Jones, D. H.; Howard, B. H. **1990**, *Biotechniques* 8, 178.
(27) Jacobson, H; Stockmayer, W. H. *J. Chem. Phys.* **1950**, *18*, 1600.
(28) Jacobson, H. *Macromolecules* **1984**, *17*, 705.
(29) Cantor, C. R.; Smith, C. L.; Mathew, M. K. *Ann. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 287.
(30) Dean, F. B.; Stasiak, A.; Koller, T.; Cozzarelli, N. R. *J. Biol. Chem.* **1985**, *260*, 4975.
(31) Krasnow, M. A.; Stasiak, A.; Spengler, S. J.; Dean, F.; Koller, T.; Cozzarelli, N. R. *Nature (London)* **1983**, *304*, 559.
(32) Kreuzer, K. N.; Cozzarelli, N. R. *Cell* **1980**, *20*, 245.
(33) Wasserman, S. A.; Cozzarelli, N. R. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 1079.
(34) Griffith, J. D.; Nash, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 3124.
(35) Wasserman, S. A.; Cozzarelli, N. R. *Science* **1986**, *232*, 951.
(36) Frank-Kamenetskii, M. D.; Vologodskii, A. V. *Usp. Fiz. Nauk.* **1981**, *134*, 641.
(37) Suematsu, K.; Ogura, H. *Bull. Chem. Soc. Jpn.* **1987**, *60*, 1177.
(38) Ma, R. I.; Kallenbach, N. R.; Sheardy, R. D.; Petrillo, M. L.; Seeman, N. C. *Nucleic Acids Res.* **1986**, *14*, 9745.
(39) Kallenbach, N. R.; Ma, R. I.; Seeman, N. C. *Nature (London)* **1983**, *305*, 829.
(40) Chen, J. H.; Kallenbach, N. R.; Seeman, N. C. *J. Am. Chem. Soc.* **1989**, *111*, 6402.
(41) Leung, Y. K.; Eichinger, B. E. *J. Chem. Phys.* **1984**, *80*, 3877.
(42) Herrmann, H. J.; Stauffer, D.; Landau, D. P. *J. Phys. A: Math. Gen.* **1983**, *16*, 1221.
(43) Mikes, J.; Dusek, K. *Macromolecules* **1982**, *15*, 93.
(44) DeBolt, L. C.; Mark, J. E. *Macromolecules* **1987**, *20*, 2369.
(45) Fyvie, T. J.; Frisch, H. L.; Semlyen, J. A.; Clarson, S. J.; Mark, J. E. *J. Polym. Sci. A* **1987**, *25*, 2503.
(46) Yuan, X. F.; Masters, A. J.; Nicholas, C. V.; Booth, C. *Makromol. Chem.* **1988**, *189*, 823.
(47) Sarmoria, C.; Valles, E.; Miller, D. R. *Macromol. Chem., Macromol. Symp.* **1986**, *2*, 69.
(48) Stanford, J. L.; Stepto, R. F. T.; Waywell, D. R. *J. Chem. Soc., Faraday Trans. 1* **1975**, *71*, 1308.
(49) Hoess, R.; Wierzbicki, A.; Abremski, K. *Gene* **1985**, *40*, 325.
(50) Marini, J. C.; Levene, S. D.; Crothers, D. M.; Englund, P. T. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 7664.
(51) Delius, H.; Gough, N. M.; Cameron, C. H.; Murray, K. *J. Virol. 1983*, *47*, 337.
(52) Hsieh, C. H.; Griffith, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 4833.
(53) Wu, H. M.; Crothers, D. M. *Biochemistry* **1986**, *25*, 3988.
(54) Ulanovsky, L.; Bodner, M.; Trifonov, E. N.; Choder, M. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 862.
(55) Hagerman, P. J. *Nature (London)* **1986**, *321*, 449.
(56) Koo, H. S.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1763.
(57) Maroun, R. C.; Olson, W. K. *Biopolymers* **1988**, *27*, 585.
(58) Levene, S. D.; Crothers, D. M. *J. Biomol. Struct. Dynam.* **1983**, *1*, 429.
(59) Dervan, P. B. *Science* **1986**, *232*, 464.
(60) Harvey, R. G.; Geacintov, N. C. *Acc. Chem. Res.* **1988**, *21*, 66.
(61) Zimmer, C.; Wahnert, U. *Prog. Biophys. Mol. Biol.* **1986**, *47*, 31.
(62) Lyubchenko, Y.; Shlyakhtenko, L.; Chernov, B.; Harrington, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 5331.