Nucleic Acids Research, 1996, Vol. 24, No. 20 **3879–3886**

Efficient random subcloning of DNA sheared in a recirculating point-sink flow system

Peter J. Oefner*, Scott P. Hunicke-Smith, Lillian Chiang[1], Fred Dietrich, John Mulligan[2] and Ronald W. Davis

Department of Biochemistry and [1]Department of Neuropharmacology, Stanford University, Stanford, CA, USA and [2]Darwin Molecular, Seattle, WA, USA

Received August 7, 1996; Accepted August 8, 1996

ABSTRACT

Based on a high-performance liquid chromatographic pump, we have built a device that allows recirculation of DNA through a 63-μm orifice with ensuing fractionation to a minimum fragment size of ~300 base pairs. Residence time of the DNA fragments in the converging flow created by a sudden contraction was found to be sufficiently long to allow extension of the DNA molecules into a highly extended conformation and, hence, breakage to occur at midpoint. In most instances, 30 passages sufficed to obtain a narrow size distribution, with >90% of the fragments lying within a 2-fold size distribution. The shear rate required to achieve breakage was found to be inversely proportional to the 1.0 power of the molecular weight. Compared with a restriction digest, up to 40% of all fragments could be cloned directly, with only marginal improvements in cloning efficiency having been observed upon prior end repair with Klenow, T4 polymerase or T4 polynucleotide kinase. Sequencing revealed a fairly random distribution of the fragments.

INTRODUCTION

Several methods have been described for random fragmentation of DNA. These methods, often used for library preparation and subcloning prior to DNA sequence analysis, include passage through the small orifice of a hypodermic needle (1–3) or of a high pressure spray atomizer (4), nebulization (5), sonic treatment (6), stirring in a blender (7,8), as well as partial digestion by restriction endonucleases (9) or treatment with DNase I in the presence of manganese ions (10). While all of these methods have been used successfully to prepare random DNA fragments for further manipulation and analysis, each has difficulties and limitations. Shearing by passage through a hypodermic needle or by stirring fails to generate fragments small enough (1–2 kb) for efficient cloning into M13 vector. Nebulization requires large volumes of DNA solution. Further, leaks in the nebulizer are common and almost unavoidable (5). Sonication, on the other hand, is difficult to reproduce, requires relatively large amounts of DNA and yields a relatively broad size distribution and, hence, a low yield of fragments useful for cloning and sequencing (5). Low cloning efficiencies have also been attributed to damage inflicted on DNA by the action of hydroxyl radicals which are known to arise as a result of thermal dissociation of water induced by pressures due to ultrasonic cavitation (11). Methods not based on the generation of hydrodynamic shear, such as the use of restriction enzymes, also have a number of disadvantages. First, some regions of DNA sequence have very few restriction sites and would be underrepresented in the resulting clone banks. Secondly, several different restriction enzymes are necessary to obtain fragments which overlap each other properly to complete the DNA sequence analysis. Thirdly, many restriction fragments are quite small and, hence, yield only very little information upon sequencing. DNase I overcomes some of these difficulties, as it cleaves with very little sequence specificity. But due to the wide size distribution of the resultant fragments, the yield of fragments having the appropriate length for cloning and subsequent sequence analysis tends to be small.

In an effort to minimize template DNA preparation tasks and simplify the preparation of libraries for shotgun sequencing, we developed a method employing an HPLC pump with a manifold valve, that allows recirculation of the DNA that is to be sheared. Major advantages of the method are a greater randomness of fragmentation sites and a >90% yield of fragments over a 2-fold size range that can be easily pre-selected by varying the flow-rate.

MATERIALS AND METHODS

Reagents

λDNA (Cat. No. 25250) and a 1 kb DNA ladder (Cat. No. 15615) were purchased from GIBCO BRL/Life Technologies (Gaithersburg, MD, USA). Cosmid DNA from human Y chromosome was isolated from the original *Escherichia coli* kindly provided by Dr Vollrath (Stanford University, Stanford, CA, USA). The shearing buffer was identical to the buffer used to store λDNA, and contained 10 mM Tris–HCl, pH 7.4, 5 mM NaCl and 0.1 mM Na_2EDTA (all from Sigma, St Louis, MO, USA). The same source of reagents was used to prepare TE-buffer (10 mM Tris–HCl, pH 8.0, 1 mM EDTA) and 5× gel loading solution (25% w/v Ficoll, 0.05% w/v bromophenol blue, 0.05% w/v xylene cyanole FF).

*To whom correspondence should be addressed at: Department of Genetics, Room M-311, Stanford University, Stanford, CA 94305-5120, USA. Email: oefner@genome.stanford.edu

Instrumentation

Shearing was carried out in a Waters Model 510 HPLC pump (Marlborough, MA), that had been equipped with a three port inlet manifold assembly (Part No. 25712, Waters). Polyether-etherketone (PEEK) tubing (1/16″ O.D.) and precut stainless steel capillary tubing (1/16″ O.D.) of inner diameters ranging from 0.0025 to 0.04″, as well as various unions and tees with the appropriate 1/16″ O.D. fittings were obtained from either Rainin Instrument Company (Woburn, MA) or Upchurch Scientific (Oak Harbor, WA). The shearing buffer was vacuum filtered through a disposable sterile bottle top filter with a 0.22-μm cellulose acetate membrane (Part No. 25970-33, Corning Glass Works, Corning, NY, USA). In addition, the buffer was continuously degassed by passing it through two channels in series of a flow-through degasser placed in-line between the reservoir and the pump inlet (DG-1310, Rainin Instrument Company). As the installation of tubings of various inner diameters results in changes in back pressure and, consequently, in the piston volume to be compressed before delivery can start, pressure-dependent deviations from the preset flow-rate do occur. Actual flow-rates for a given combination of tubings were determined by collecting the solvent in a volumetric flask and noting the time. Subsequently, the volume flow/min was calculated and given in the figure legends. DNA concentrations were determined by means of a Spectronic 1001 split-beam spectrophotometer (Bausch & Lomb Inc., Rochester, NY, USA).

Hydrodynamic shear breakage of DNA in recirculating mode

Buffered solutions of 0.5–80 μg DNA were loaded with a Hamilton microliter syringe (Reno, NE, USA) into the 50- or 100-μl sample loop of a seven-port sample injection valve (Rheodyne, Model 7725i, Cotati, CA, USA). Following the switching of the valve from LOAD to INJECT, with the manifold valve in the RECYCLE position, mobile phase was flushed through the loop for 30 s at a flow-rate of 0.5 ml/min. Then it was returned to the LOAD position, in order to keep the internal volume of the device as small as possible. Finally, the flow-rate was adjusted to obtain fragments of the desired size range. After a given period of time, the manifold valve was switched to the COLLECT position and 0.7–1.0 ml of sheared DNA was collected. The eluate was precipitated with 0.1 vol 3 M sodium acetate, pH 7.0 and 1 vol isopropanol. The sample was mixed, incubated at –30°C for at least 90 min and spun at 16 000 *g* for 30 min. The pellet was washed twice with 75% ethanol, dried under vacuum and, finally, resuspended in 16 μl of TE buffer and 4 μl of 5× sample loading solution. Between shearing runs of DNA derived from the same source, the device was flushed with several milliliters of shearing buffer, the manifold valve being in WASTE position. Between the shearing of different DNA samples, the device was rinsed first with 0.2 M HCl followed by 0.2 M NaOH to hydrolyze any remaining DNA in the device and, hence, to avoid carry-over. Care has to be taken to completely flush out the sodium hydroxide solution with shearing buffer prior to the next injection. Otherwise, denaturation of the DNA will result in a second band of fragments exhibiting roughly half the expected size. This band contains single-stranded DNA molecules that break twice as easily as double-stranded DNA (12,13).

Agarose electrophoresis

Sheared DNA samples were separated on a 1% agarose slab gel (SeaKem GTG Agarose, FMC BioProducts, Rockland, ME) in 45 mM Tris-borate buffer, pH 8.3, containing 1 mM EDTA. Separations were carried out in a horizontal agarose submarine unit (Model MGU-200T, C.B.S. Scientific Co., Del Mar, CA, USA) at a constant voltage of ~115 V. Size of the sheared DNA fragments was computed by digitizing agarose slab gels with the IS-1000 Digital Imaging System from Alpha Innotech Corp., San Leandro, CA, USA. The image was calibrated with DNA fragments of known size, and then a cubic spline interpolation function was used to map digitized coordinates to fragment length in base pairs.

Evaluation of cloning efficiency

Cloning efficiency was determined by two methods: as the ratio of tetracycline-sensitive (tet^s) to ampicillin-resistant (amp^r) transformants arising when DNA fragments were blunt-end ligated into the unique *Eco*RV site within the tet^r gene of pBR322; or alternatively, as the efficiency with which amp^r transformants were obtained when increasing quantities of the DNA fragments were ligated to alkaline phosphatase-treated linearized pBR322.

Comparisons were made between the cloning efficiency of untreated HPLC-sheared λDNA fragments, those treated with Klenow [DNA fragments at 0.2 μM, in 50 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 50 μg/ml BSA and 25 μM each dNTP with 5 U Klenow (United States Biochemicals, Cleveland, OH) in a 20 μl volume for 15 min at 30°C], T4 DNA polymerase [DNA fragments at 0.2 μM, in 33 mM Tris-acetate, pH 7.9, 66 mM potassium acetate, 10 mM magnesium acetate, 0.5 mM DTT, 0.1 mg/ml BSA and 25 μM each dNTP with 6 U T4 DNA polymerase (United States Biochemicals) for 5 min at 37°C], or T4 polynucleotide kinase (see below), and λDNA fragments obtained from *Hinc*II digestion (New England Biolabs, Beverly, MA). λDNA fragments obtained from *Pst*I and/or *Eco*47I digestion (New England Biolabs) with/without Klenow or T4 DNA polymerase treatment were concurrently cloned as controls for efficient end-repair. T4 polynucleotide kinase (5–10 U/20 μl reaction volume; Life Technologies, New England Biolabs or United States Biochemicals) was utilized to treat HPLC-sheared λDNA fragments (for 30 min at 37°C) under forward reaction conditions (0.2 μM DNA fragments, 50 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 1.5 mM spermidine, 0.1 mM EDTA and 0.1 mM ATP), exchange reaction conditions (0.2 μM DNA fragments, 50 mM imidazole–HCl, pH 6.4, 12 mM $MgCl_2$, 0.5 μM ATP, 0.3 mM ADP and 15 mM β-mercaptoethanol), and optimal 3′-phosphatase conditions (0.2 μM DNA fragments, 0.1 M MES, pH 6.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 50 μg/ml BSA); efficiency of incorporation under the first two conditions was monitored with [γ^{32}P]ATP (NEN Dupont).

After *Eco*RV digestion (New England Biolabs), alkaline phosphatase-treated vector (10 μg linearized pBR322 in 400 μl 10 mM Tris–HCl, pH 8.0, incubated for 30 min at 37°C with 0.04 U alkaline phosphatase from Boehringer Mannheim) was phenol–chloroform extracted three times, chloroform extracted once, and ethanol precipitated. The concentrations of treated/untreated fragments and phosphatased/phosphorylated vector were determined immediately before use in ligation by spectrophotometry (estimating 50 μg/ml double-stranded DNA for an OD_{260} of 1). Ligations were peformed overnight at 16°C in a 20 μl volume with 10.4 nM vector and an appropriate amount of fragment (62.1

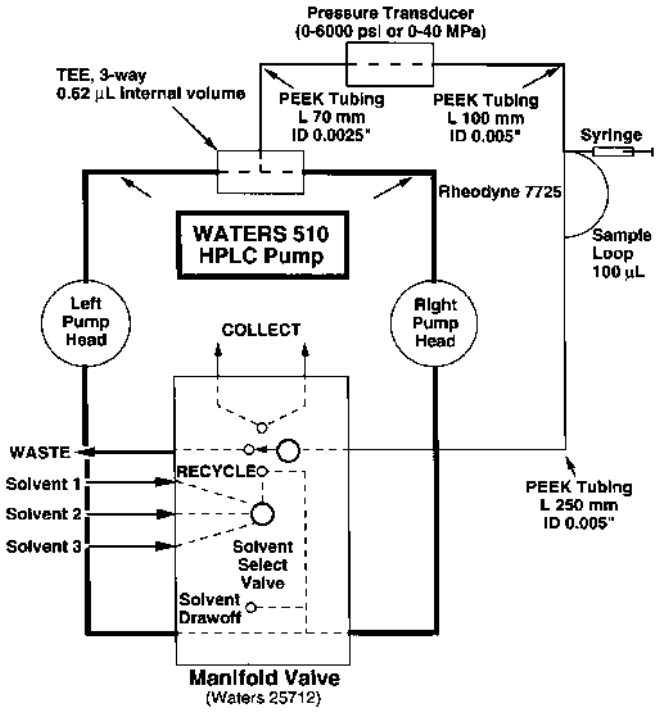

Figure 1. Scheme of the apparatus employed for shearing DNA. Arrows indicate direction of flow.

nM when cloning into phosphorylated vector; 0–270 nM in incremental amounts when cloning into phosphatased vector), in 50 mM Tris–HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP and 50 µg/ml BSA, using 400 U T4 DNA ligase (New England Biolabs). After heat inactivation (10 min at 70°C), 3 µl of each ligation mix was transformed into Subcloning Efficiency DH5α competent cells (Life Technologies). Transformants were selected on LB plates containing 50 µg/ml ampicillin; tetracyline-sensitive clones were screened by patching on LB plates containing 20 µg/ml tetracycline. Transformation of vector alone ligations determined background in these assays.

The fraction of clonable fragments in a given HPLC-sheared population (with/without additional end-repair) was estimated by determining the percent recombinants (tet^s/amp^r transformants) relative to the percent recombinants obtained for *Hinc*II-digested λDNA (chosen because *Hinc*II digestion produces 35 blunt-ended fragments per λDNA ranging in size from 228 to 4755 bp and HPLC-sheared fragments averaging 1000 bp in length results in ~50 fragments per λDNA). Alternatively, the number of transformants obtained/µg vector was plotted against nM fragment ends (estimated for HPLC-sheared fragment by assuming 50 fragments per λDNA); the fraction of clonable fragments in comparison with *Hinc*II-digested λDNA was then taken as the ratio of slopes at the linear portion of the curves.

Determination of the DNA sequence adjacent to the cleavage site

For determining the sequence of the fragment ends, 0.4–5 µg of cosmid or P1 cloned human DNA was sheared in 10 mM Tris–HCl, pH 8.0, 5 mM NaCl and 1 mM EDTA. After 8 min of shearing, 1 ml of sheared fragments was collected into a Centricon 30 (Amicon, Beverly, MA) concentrator. The samples were concentrated to 50 µl by centrifugal dialysis according to the manufacturer's instructions. One ml of TE (10 mM Tris, 1 mM EDTA) was added to each centricon and the centrifugal concentration repeated. Vector DNA was prepared by digestion of M13mp19 (New England Biolabs) with *Sma*I (New England Biolabs) and treatment with Calf Intestinal Alkaline Phosphatase according to the manufacturer's instructions (Boehringer Mannheim). In some cases the sample was cloned directly; for later libraries the sheared DNA was first cleaned up by gel purification and isolated using a Geneclean II kit (Bio101, La Jolla, CA). Ligations were carried out in 20 µl with 20 ng of vector and 10–100 ng of sheared DNA. The ligations were incubated for 10–16 h at 16°C with T4 DNA ligase (New England Biolabs) in 50 mM Tris–HCl, pH 7.8, 10 mM $MgCl_2$, 10 mM DTT, 500 µM ATP, 25 µg/ml BSA, at 65°C for 5 min, electroporated into DH10B cells and then plated on DH5α F′ plating cells (both from Gibco BRL Life Technologies) prepared according to Maniatis *et al.* (9). A total of 1000–40 000 recombinant clones were produced in each library. Without gel purification, the libraries contained 5–10% of short inserts (<400 bp, in some cases as short as 80 bp). From each cosmid or P1 clone, between 1000 and 3000 clones were sequenced using the dye-primer sequencing chemistry from the Applied Biosystems Division of Perkin-Elmer (Foster City, CA, USA).

Determination of fragment distribution

For DNA subfragment cloning, adapters with a CAAAC overhang were either ligated directly to the sheared DNA fragments overnight at 16°C with T4 DNA ligase, or after end-repair of the DNA fragments with T4 DNA polymerase at 11°C for 60 min. Following ligation, the DNA subfragments were separated from adapter dimers by means of slab gel electrophoresis on a low-melting 0.8% agarose gel. Subsequently, they were ligated into M13mp18 vector at 16°C overnight with T4 DNA ligase and transformed into NM 522 *E.coli* cells (Promega, Madison, WI, USA). Subclone DNAs were prepared by lysis of the phages in ethanol–butanol (5:2, v/v), purification of the released DNA with Cleanascite (Affinity Technologies, New Brunswick, NJ, USA), followed by ethanol precipitation. Subclones were sequenced on a Catalyst 800 Cycle Sequencer (Perkin-Elmer, Applied Biosystems Division) using a linear amplification method with fluorescent dye-primers, followed by electrophoresis on a 4.75% acrylamide gel run on an Applied Biosystems Model 373 Stretch Sequencer.

RESULTS AND DISCUSSION

General

Figure 1 shows a scheme of the apparatus used for shearing DNA in a recirculating point-sink flow system. The device had been developed originally for the preparative chromatographic separation and isolation of an intermediate in the total synthesis of vitamin B12 (14). The corner stone of the apparatus is the manifold valve, which allows to direct the solvent to the WASTE, COLLECT or RECYCLE position respectively. In the RECYCLE mode, the pump does not draw any solvent from the solvent reservoir. Instead, solvent is recirculated through the system. The dead volume of the system in the RECYCLE mode is ~1 ml as determined by the photometric measurement of DNA dilution occurring over time. These measurements also revealed no

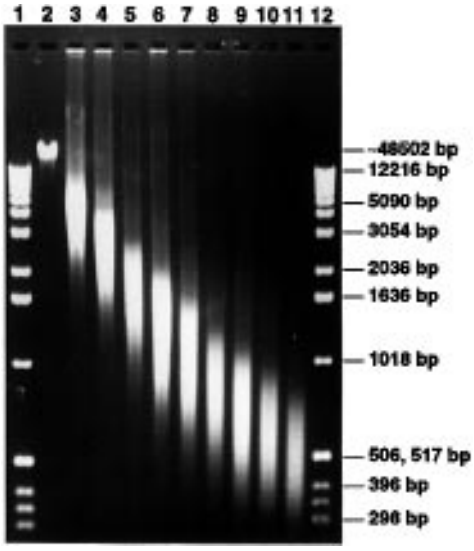

Figure 2. Agarose slab gel electrophoresis of λDNA sheared at various flow-rates by means of the apparatus depicted in Figure 1, which had been equipped with the 0.62-µl stainless steel tee. Lanes 1 and 12: 1 kb DNA ladder (0.5 µg each); lane 2: unsheared λDNA (~0.1 µg); lanes 3–11: λDNA sheared at flow-rates of 1.35, 2.3, 3.2, 4.1, 5.0, 5.95, 6.9, 7.7 and 8.6 ml/min respectively, for 8 min each. Amount of λDNA sheared: 4.08 µg each.

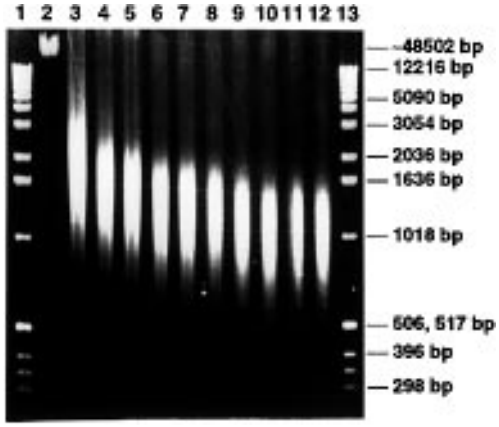

Figure 3. Agarose slab gel electrophoresis of λDNA sheared at a flow-rate of 5.0 ml/min for different periods of time by means of the apparatus used in Figure 2. Lanes 1 and 13: 1 kb DNA ladder (0.5 µg each); lane 2: unsheared λDNA (~0.1 µg); lanes 3–12: λDNA (4.08 µg) sheared at a flow-rate of 5.0 ml/min for 0.33, 0.66, 1, 2, 3, 4, 6, 8, 10 and 15 min respectively.

significant loss of DNA during shearing, independently from the set flow-rate.

Figure 2 shows the slab gel electrophoretic analysis of shear experiments carried out at various flow-rates ranging from 1.4 to 8.6 ml/min through the aforementioned device, which was equipped with a 0.62-µl stainless steel tee, to the outlet of which a 4.8-cm PEEK tubing of 0.0025″ I.D. had been attached. It is evident that, with increasing flow-rates, λDNA could be degraded to increasingly smaller fragments. At the maximum flow-rate of 8.6 ml/min, fragments ranged in size from ~300 to 600 bp.

From Figure 3 it is obvious that the final size distribution at a flow-rate of 5 ml/min was achieved within 6 min, or 30 passes through the instrument. This also explains the broader size distribution observed at lower flow-rates in Figure 2, as the number of passes through the device was too small. Shearing times >6 min increased the percentage of fragments within a range of 750–1500 bp only from ~91 to 93.5% as determined by densitometry, hence indicating that after a certain number of passages no further changes in the product can be discerned upon continued cycling. This final size distribution is in agreement with the hypothesis that in a given shear gradient only molecules above a certain size are cleaved, while the smaller ones escape fragmentation. Therefore, the narrowest possible distribution will have a 2-fold spread in size assuming that cleavage occurs in the middle of a molecule; that is, the highest molecular-weight material should have a little less than twice the lowest molecular weight. This appears to be nearly the case in the present study. This was not only shown by densitometry of sheared DNA solutions run on agarose slab gels but also by PCR amplification of the cloned fragments with subsequent size determination by slab gel electrophoresis. Both methods corroborated that routinely >90% of all sheared fragments are located within a 2-fold molecular weight range. Further, the size distribution attained at a given flow-rate is not dependent on the source of DNA; this includes restriction fragments, cosmids, BACs, YACs and even whole genomic DNA. Therefore, once the instrument has been calibrated, which is necessary due to the variability (±0.001″ according to the manufacturer) of the inner diameter of the 0.0025″ tubing attached to the outlet of the tee, identical size distributions can be expected for any DNA sheared thereafter. Finally, up to DNA concentrations of 40 µg/ml, no significant self-protection effects were observed that would result in larger fragments than expected from the applied shear stress.

While DNA library preparation for shotgun sequencing usually requires DNA fragments in the size range of 1–4 kb, there are other applications that will benefit from the generation of larger, but still very well size-defined DNA fragments. The latter can be readily accomplished by the present method by shearing DNA through the same apparatus at lower flow-rates (Fig. 4).

Effect of ionic strength on shear degradation

In contrast with studies by Bowman and Davidson (15) as well as Pyeritz *et al.* (12), we were able to observe that an increase in ionic strength had a significant effect on the breakage rate. A higher shear rate was required to obtain identical breakage rates when shearing was carried out in 200 mM instead of 5 mM sodium chloride. This accords well with the observation of Yew and Davidson that the rate of breakage at a given shear stress increases with decreasing ionic strength, approximately as $[Na^+]^{-1.6}$ (16). It is also in agreement with the well established effect of ionic strength on intrinsic viscosity of a DNA solution, which in the case of T7 DNA has been reported to increase by a factor of ~1.5 when the salt concentration is decreased from 0.2 to 0.005 M (17). Hence, it can be concluded that DNA attains a more coiled conformation the higher the salt concentration. Consequently, a higher shear stress is required to obtain fragments of identical size distribution with an increase of the concentration of sodium chloride in the shearing buffer from 0.005 to 0.2 M NaCl (Fig. 5).

It has been reported that the likelihood that a shear-broken fragment will contain a single-chain terminus increases with decreasing salt concentration (12). At a temperature of 37°C, for

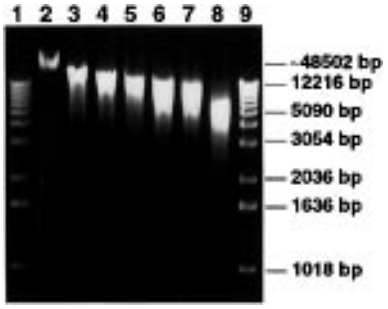

Figure 4. Agarose slab gel electrophoresis of λDNA sheared at various flow-rates by means of the apparatus depicted in Figure 1. Lanes 1 and 9: 1 kb DNA ladder (0.4 µg each); lane 2: unsheared λDNA (~0.1 µg); lanes 3–8: λDNA (2.04 µg) sheared at flow-rate of 0.1, 0.2, 0.3, 0.4, 0.5 and 1.0 ml/min for 30, 15, 10, 10, 10 and 8 min respectively.

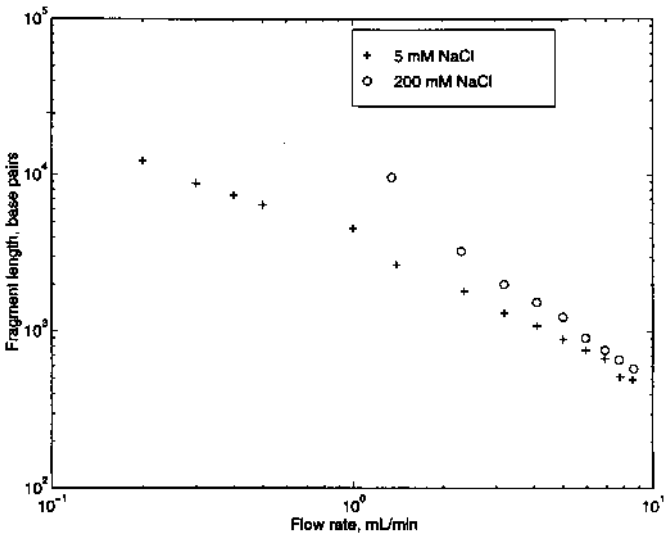

Figure 5. Mean fragment length in base pairs is plotted as a function of flow rate for two salt concentrations.

instance, the percentage of the single-chain termini increased from 20 to 71.5% when the sodium chloride concentration was lowered from 2.5 to 0.01 M respectively. No attempt was made in the present study to determine the presence of single-chain termini. However, no improvement in clonability of sheared fragment was observed upon increasing the concentration of sodium chloride from 5 to 200 mM, although it cannot be dismissed that any beneficial effect of the higher ionic strength buffer was missed due to the well known variability of cloning experiments, with decreases in the percentage of single-chain termini having been reported to be <20% upon increasing the concentration of sodium chloride from 10 to 100 mM (12). Even if higher ionic strength results in a significant reduction of single-chain ends, it is rather unlikely that a significant improvement in cloning efficiency will be observed. In the present study, no consistent improvement has been observed following treatment with Klenow, T4 DNA polymerase or T4 polynucleotide kinase. Furthermore, the percentage of single-chain termini may be no real indicator of the clonability of sheared fragments, as the same study (12), which had reported such high percentages of single-chain termini following passage of DNA through a No. 28 hypodermic needle, did not detect any single-chain ends in sonicated DNA fragments. The latter result would have indicated that it might be possible to clone sonicated fragments without any prior end-repair. However, the contrary was reported by Deininger (6). His attempts to directly clone sonicated DNA had not produced any clones containing inserted DNA. In the present study, on the other hand, 20–40% of HPLC-sheared λDNA fragments could be cloned in the absence of additional end-repair.

The 2-fold size distribution of the DNA fragments also suggests that shear breakage carried out under the conditions described does not introduce substantial amounts of single-chain nicks, as they constitute preferential sites of shear breakage that would result in substantially smaller fragments and a broader size distribution. That single-stranded DNA breaks twice as readily as double-stranded DNA was also confirmed by shearing single-stranded viral DNA, which yielded a 2-fold distribution of fragments half of the size expected from the shear breakage of double-stranded DNA (data not shown). This is also in accordance with the findings by Odell and Taylor, who used a four-roll mill for their experiments (13).

Future experiments will show whether transition metals such as lanthanum III ions are able to catalyze the reaction as demonstrated previously; Adam and Zimm have observed a 10-fold increase in breakage rate in the presence of 50 µM La^{3+} at pH 7.5 (18). This will also confirm whether the breakage mechanism is indeed a phosphate ester hydrolysis, wherein a hydroxyl ion attacks the phosphorus, forming a pentacoordinate intermediate, with a nearby water molecule transferring a proton to the departing group in a concerted manner as it comes off the phosphorus. The present study provides at least indirect support through the observation that a substantial number of fragments can be ligated and cloned into a M13 vector without prior end-repair with T4 polymerase, while this has been reported to be impossible after sonic degradation (6), wherein ~90% of cleavage has been determined to occur by C–O bond rupture (19).

Kinetics of breakage

The physics of polymer stretching and breakage have been the subject of numerous reviews (20–22). While details of the kinetics have not been resolved, the dynamics of breakage are driven by the hydrodynamic drag force on the molecule and the spring force internal to the molecule. All current models of long chain molecules in flow assume the drag on the molecule is proportional to the local relative velocity of the surrounding fluid. While DNA molecules are in a coiled state in the absence of flow or in sufficiently weak flows, their central portions disentangle as soon as a critical velocity gradient is reached, and become more and more straightened out along the direction of flow at the expense of the two end portions which remain curled (21). This non-uniform stretching of polymers, with a taut central portion and two randomly coiled end portions, occurs for the simple reason that the viscous pull due to extensional flow is not uniform along the length of the chain, but roughly parabolic, being zero at the ends and maximum in the middle (23). This also explains the observed 2-fold size distribution of our data (for multiple breakage passes) and that of Bowman and Davidson (15). The latter reported a mean single molecule break at half length with a standard deviation of 12.5%. Thus if a very long molecule were subjected to repeated breaks, each in the middle ±12.5%, the

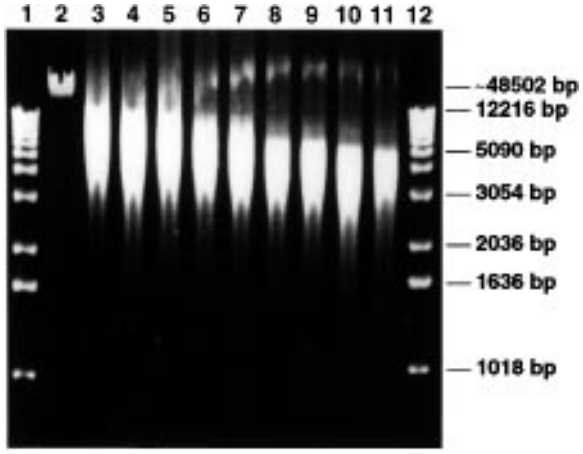

Figure 6. Agarose slab gel electrophoresis of λDNA sheared at various flow-rates by means of the apparatus depicted in Figure 1, which had been equipped with the 0.62-µl stainless steel tee, but with the 0.0025″ I.D. PEEK tubing at its outlet having been replaced with a 0.005″ I.D. stainless steel tubing. Lanes 1 and 12: 1 kb DNA ladder (0.5 µg each); lane 2: unsheared λDNA (~0.1 µg); lanes 3–11: λDNA (4.08 µg) sheared at flow-rates of 1.4, 2.4, 3.25, 4.2, 5.15, 6.1, 7.05, 8.05 and 9.0 ml/min respectively, for 8 min each.

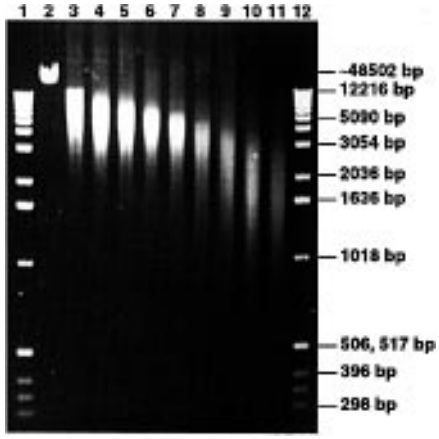

Figure 7. Agarose slab gel electrophoresis of λDNA sheared at various flow-rates by means of the apparatus depicted in Figure 1, which had been equipped with the 0.62-µl stainless steel Tee, but with the 0.0025″ I.D. PEEK tubing having been relocated from the outlet of the Tee to that of the pressure transducer. Lanes 1 and 12: 1 Kb DNA ladder (0.5 µg each); lane 2: unsheared λDNA (~0.1 µg); lanes 3–11: λDNA (4.08 µg) sheared at flow-rates of 1.4, 2.4, 3.25, 4.2, 5.15, 6.1, 7.05, 8.05 and 9.0 ml/min respectively, for 8 min each.

limiting result would be a population of molecules having a 2-fold size distribution.

Breakage occurs when the cumulative hydrodynamic drag creates enough tension in the molecule to first unravel and then break or catalyze the breakage of the chemical bonds. This means that it is strictly velocity gradients, or strain rate, in the direction of the molecule's backbone which causes sufficient differential drag on the molecule to break it. Further, given repeated exposure of sufficient duration, all molecules will be broken by the maximum strain rate in a flow and this in turn will define the final median size. One would therefore like to select a fluid flow which subjects all molecules to an identical strain rate in the direction of bulk fluid flow for an extended period of time. The ideal flow is plane stagnation point flow, such as in a four-roller mill or impinging jet. While Odell and Taylor (13) argue that it is only possible to have true mid-point breakage in these flows due to the long residence time of the DNA in strain, the point-sink flow used in the present study also appears to allow sufficient extension so that breakage product populations are nearly the same. In contrast, the mechanism chosen by Bowman and Davidson (shear breakage in a capillary tube) has a strain rate varying from maximum at the tube wall to zero at the tube centerline and causes molecules to tumble with respect to the maximum strain direction. This tends to broaden the distribution of DNA fragments.

Since the contraction from large to small ID tubing (0.015/0.0025 = 6) causes such a significant change in cross-sectional area and the fluid is essentially incompressible, the governing fluid mechanics equations may be solved to determine the local strain rate. By neglecting effects of fluid viscosity and the presence of the DNA on the fluid flow, which is justified in the present case both due to the large values of Reynolds number ranging from 680 at 2 ml/min to 2700 at 8 ml/min (25), and the low concentrations of DNA, the exact theoretical solution for fluid motion for this point-sink flow gives a strain rate of:

$$dv / dr = Q / 2\pi r^3 \quad \mathbf{1}$$

where v is the fluid velocity, Q is the volume rate, and r is the radial distance from the sink origin. Since the model breaks down as r approaches zero, the orifice radius (0.00125″) was taken to compute dv/dr for correlation with DNA fragment length. The point of this model is to provide a reasonable representation of actual maximum strain rate for the purposes of correlation with terminal fragment size. The data only show that the terminal fragment size is well correlated with the kinematic quantity, strain rate; they do not necessarily imply that the physical model of the system is correct. As Figure 5 shows, final DNA fragment size is indeed a reproducible function of volume flow rate Q. These data and previous studies suggest that the strain rate (dv/dr) required for breakage is proportional to M^{α}, where M is the fragment molecular weight and α varies with solvent properties and flow geometry. Our data find $\alpha \sim -1$ which agrees well with the values of –0.95 to –1.1 reported by Odell *et al.* (24) for similar flow geometries but much lower strain rates (10^3/s versus 10^6/s) and different polymers (bis-phenol-A polycarbonate and polystyrene).

As evident from Figure 6, only degradation to a size of ~5000 bp was obtained when the 0.0025″ I.D. tubing at the tee outlet had been replaced with a 0.005″ I.D. tubing. As the model above suggests, this slight change in diameter makes a significant change in strain rate, and thus in final fragment length. It is also interesting to note that an increase in flow-rate did not primarily affect the shear rate or minimum fragment size obtainable, but rather resulted in a more narrow size distribution.

Finally, it was observed that replacement of the 0.0025″ ID tubing with a ruby orifice of the same inner diameter but a length of only 100 µm did not yield the same rate of breakage at a given flow-rate. Therefore, it is assumed that the fluid flow immediately beyond the contraction causes a further reduction in the cross-sectional area, hence contributing significantly to the shear stress and, ultimately, the small fragment size obtained.

Flow-preconditioning

It was also noticed that by moving the 0.0025″ ID tubing farther downstream of the tee outlet, the resulting mean terminal fragment length was significantly larger than was previously the case at the same flow-rate (Fig. 7). The only apparent change to

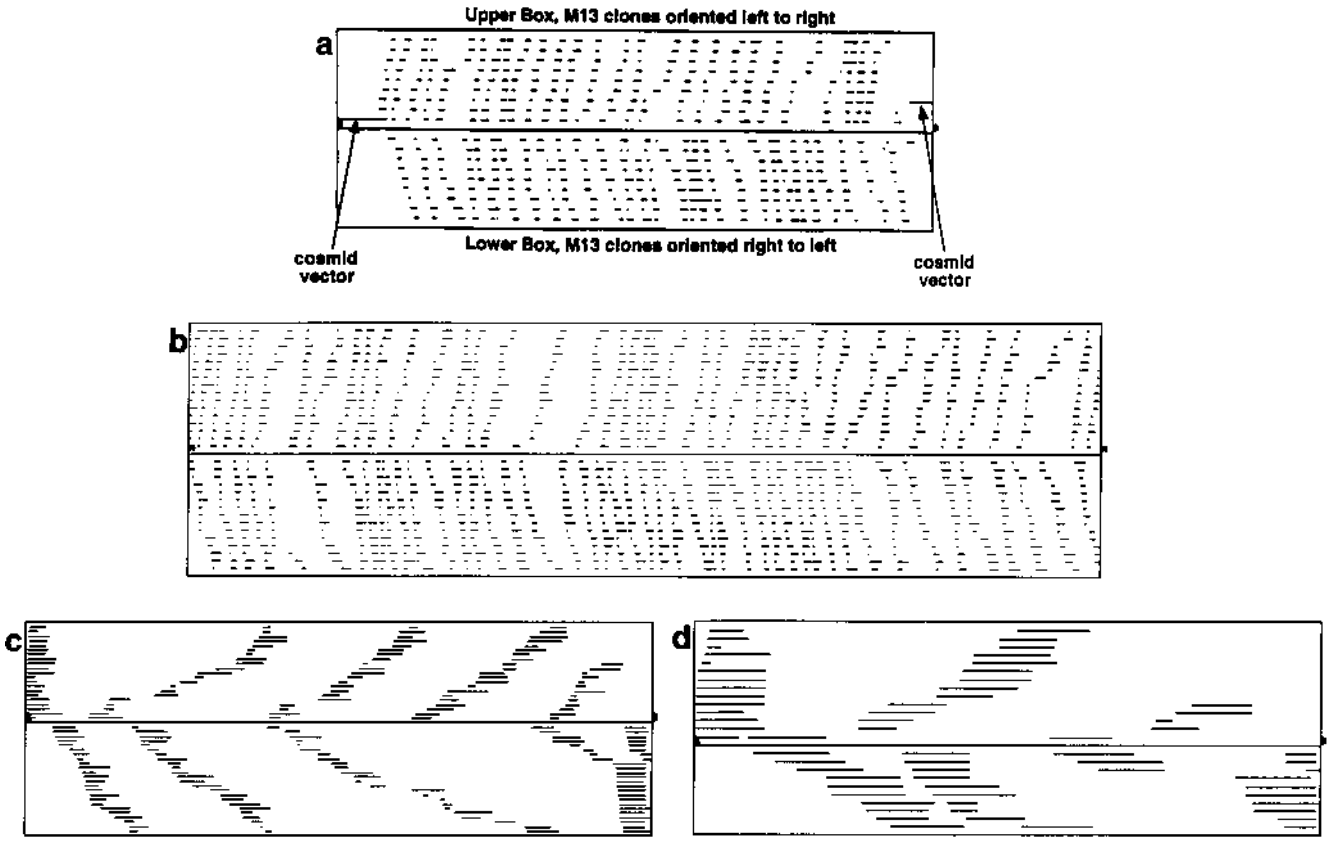

Figure 8. Distribution of (**a**) 1258 subclones in a human Y chromosome cosmid, and (**b**) 206 subclones in a 7.68-kb *Nhe*I restriction fragments of *Chlamydia trachomatis* respectively, obtained after hydrodynamic shear breakage and subcloning in M13.

the flow kinetics is the distance between the 90° flow bend in the tee and the region of high shear. In order to gain more insight into the observation that the proximity of the shearing contraction to the tee affected the mean size of the sheared DNA, we constructed scale fluid models of the tee and passageway, matching the critical fluid dynamic scaling parameter, namely the Reynolds number. By dye injection and qualitative observation, we noted that the fully developed velocity profile of the fluid entering the tee was distorted into a pair of counter-rotating vorticies, with axes oriented in the bulk flow direction, as the flow turned the corner and entered the contraction. We speculate, based on this observation, that the vorticity associated with fully developed tube flow (causing tumbling of the DNA molecules) has been restructured so that DNA molecules are more likely to be aligned with the direction of maximum extension.

Statistical analysis of break points

Comparison of the expected nucleotide composition with that observed for 576 fragment ends generated from shearing clones of human DNA and cloning the fragments directly without any further enzymatic treatment into a *Sma*I cut, phosphatased vector, revealed that Gs and Cs are overrepresented ($\chi_{[1]} = 22.7$, $P < 0.001$). Looking at the base pairs flanking the cleavage sites, rather than base composition, AT/GC and CG/GC were overrepresented [$\chi_{[1]}$ of 11.7 and 13.22 ($P < 0.001$) respectively]. On the other hand, AA/TT ($\chi_{[1]} = 12.5$, $P < 0.001$) and TA/AT ($\chi_{[1]} = 8.17$, $P < 0.005$) were found to be underrepresented. The reason for this bias is not completely clear. It appears that GC-rich DNA cleaves more readily. However, the data may also reflect reduced efficiency with which fragments having terminal thymines and adenines can be cloned. Effects of ultrasound on nucleic acid bases has shown that sonochemical changes induced mainly by hydroxy radicals decrease in the order thymine > cytosine > guanine > adenine (11). It can not be excluded that in the point-sink flow system used in the present study, thermal and pressure induced dissociation of water does occur. The occurrence of radicals would also explain why the percentage of clonable fragments can not be increased significantly by enzymatic end repair.

However, from the schematic representation of M13mp18 subclones obtained from a sheared cosmid containing a human Y chromosome insert, it is obvious that the distribution of clones is sufficiently random to ensure complete coverage (Fig. 8a). Only two ~300-bp regions of the ~41-kb cosmid were covered only by one M13 clone on one strand. The rest of the cosmid sequence was covered by multiple clones on both strands.

In the case of the hydrodynamic shear breakage of restriction fragments, terminal regions tended to be overrepresented, presumably due to the 100% efficiency with which the enzymatically cleaved termini can be ligated (Fig. 8b). Although this indicates that shear breakage leads to the generation of fragment ends that can not be repaired successfully, there were still enough fragments obtained to determine the complete sequence of the restriction fragment by a total shotgun approach, i.e. there was no need for filling gaps using a nonrandom approach.

Future progress

Since the method employs chromatographic equipment, it lends itself easily to automation, using an electric valve for automating the flushing of lines between shearing runs, an autosampler for injecting the samples, and a fraction collector for collecting the fragments. Work is also underway to develop a simpler dedicated system for DNA shearing based on a syringe pump, that should considerably reduce the cost of equipment.

ACKNOWLEDGEMENTS

We are grateful to Dr Richard Hyman for the preparation of DNA libraries, and to Drs Richard Stephens and Sue Kalman for provision of the restriction fragment from *Chlamydia trachomatis.* We also thank Dr George Homsy for helpful fluid mechanics discussions. This work was supported by a grant from NIH (1P01-HG00205-05).

REFERENCES

1 Davison, P. F. (1959) *Proc. Natl. Acad. Sci. USA* **45**, 1560–1568.
2 Davison, P. F. (1960) *Nature* **185**, 918–920.
3 Schriefer, L. A., Gebauer, B. K., Qiu, L. Q. Q., Waterston, R. H. and Wilson R. K. (1990) *Nucleic Acids Res.*, **18**, 7455–7456.
4 Cavalieri, L. F. and Rosenberg, B. H. (1959) *J. Am. Chem. Soc.* **81**, 5136–5139.
5 Bodenteich, A., Chissoe, S., Wang, Y.-F. and Roe, B. A. (1994) In Adams, M. D., Fields, C. and Venter, J. C. (eds) *Automated DNA Sequencing and Analysis*. Academic Press, San Diego, CA.
6 Deininger, P. L. (1983) *Anal. Biochem.* **129**, 216–223.
7 Hershey, A. D. and Burgi, E. (1960) *J. Mol. Biol.* **2**, 143–152.
8 Rosenberg, H. S. and Bendich, A. (1960) *J. Am. Chem. Soc.* **82**, 3198–3201.
9 Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982) *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
10 Anderson, S. (1981) *Nucleic Acids Res.* **9**, 3015–3027.
11 McKee, J. R., Christman, C. L., O'Brien, W. D. and Wang, S. Y. (1977) *Biochemistry* **16**, 4651–4654.
12 Pyeritz, R. E., Schlegel, R. A. and Thomas, C. A., Jr (1972) *Biochim. Biophys. Acta* **272**, 504–509.
13 Odell, J. A. and Taylor, M. A. (1994) *Biopolymers* **34**, 1483–1493.
14 Woodward, R. B. (1973) *Pure Appl. Chem.* **33**, 145–177.
15 Bowman, R. D. and Davidson, N. (1972) *Biopolymers* **11**, 2601–2624.
16 Yew, F. F. H. and Davidson, N. (1968) *Biopolymers* **6**, 659–679.
17 Rosenberg, A. H. and Studier, F. W. (1969) *Biopolymers* **7**, 765–774.
18 Adam, R. E. and Zimm, B. H. (1977) *Nucleic Acids Res.* **4**, 1513–1537.
19 Richards, O. C. and Boyer, P. D. (1965) *J. Mol. Biol.* **11**, 327–340.
20 Reese, H. R. and Zimm, B. H. (1990) *J. Chem. Phys.* **92**, 2650–2662.
21 Larson, R. G. (1990) *Rheol. Acta* **29**, 371–384.
22 Wiest, J. M., Wedgewood, L. E. and Bird, R. B. (1989) *J. Chem. Phys.* **90**, 587–594.
23 Frenkel, J. (1944) *Acta Physiochim. URSS* **19**, 51–76.
24 Odell, J.A., Keller, A. and Muller, A. J. (1992) *Colloid & Polymer Sci.* **270**, 307–324.
25 Ryskin, G. (1987) *J. Fluid Mech.* **178**, 423–440.