Single molecule detection of direct, homologous DNA/DNA pairing

C. Danilowicz[a], C. H. Lee[a], K. Kim[b], K. Hatch[a], V. W. Coljee[a], N. Kleckner[b], M. Prentiss[a,1]

[a] Department of Physics, Harvard University, Cambridge, MA 02138, USA.
[b] Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

[1] Corresponding author.

e-mail : prentiss@fas.harvard.edu

tel: (617) 495-4483

Classification**:** BIOLOGICAL SCIENCES- Biophysics and Computational Biology

Summary

We use a parallel single molecule assay that utilizes magnetic tweezers to demonstrate homologous pairing of two double-stranded (ds) DNA molecules in the absence of proteins, divalent metal ions, crowding agents or free DNA ends. Under physiological conditions of temperature and monovalent salt, pairing is quite accurate and rapid, even at DNA molecule concentrations that are orders of magnitude below those found *in vivo* and in the presence of a large excess of nonspecific competitor DNA. Crowding agents further increase the reaction rate. Pairing is readily detected between regions of homology of 2kb or more. The detected pairs, of all lengths, are stable not only against thermal forces but to shear forces up to 10pN. The rate of pairing increases monotonically with the square of the DNA concentration, the length of homology, and the concentration of either K^+ or Na^+. Pairing increases with temperature up to 40°C and then decreases. Several lines of evidence exclude Watson-Crick base pairing as the basis for pairing. These results strongly suggest that direct recognition of homology between chemically intact B-DNA molecules should be possible *in vivo*. Indeed, the robustness of the observed signal raises the possibility that pairing might be the "default" option such that, in general, specific features must be present to limit pairing to particular desired situations. The possibility of protein-independent homologous pairing of intact dsDNA has been predicted theoretically, with more than one pairing mechanism proposed. Further studies will be required to determine whether existing theories fit the sequence length, temperature and salt dependencies described in this work.

Introduction

Pairing of homologous DNA/chromosome regions is a central feature of many biologically important processes. Recombinational double-strand break repair and programmed homologous recombination during meiosis all involve complex series of biochemical reactions in which single-stranded (ss) DNA plays a prominent role. Homology recognition and pairing reactions mediated by bacterial RecA protein and its archael and eukaryotic relatives are initiated by loading of multiple protein molecules onto a ssDNA segment; the resulting protein/ssDNA filament then searches for a homologous region of duplex DNA (1). In another type of reaction, "single-strand annealing" (SSA), two complementary ssDNA regions directly anneal to one another as mediated by other types of proteins (2). The basic natures of homology recognition for these reactions are largely understood.

There also exist homologous pairing reactions that, as far as can be discerned, involve interactions between chromosomal regions whose DNAs are chemically intact double-stranded (ds) DNA. In contrast to recombination-related processes, the basis for homology recognition in these "recombination-independent" pairing processes remains essentially unknown.

The original and best-studied case of such "recombination-independent pairing" occurs in the fruit fly *Drosophila melanogaster* (3). In this organism, the maternal and paternal versions of each chromosome (i.e. "homologs") are juxtaposed all along their lengths from the 12th-13th divisions of the developing embryo onward. This "somatic pairing" is then further reinforced during the specialized program of meiosis. Recombination-independent pairing is the rule for Drosophila male meiosis and is also present during female meiosis, in addition to a meiosis-specific recombination-dependent

mechanism. Several lines of evidence suggest that, in all of these situations, pairing involves interactions that occur independently at many different positions along the chromosomes, a mode that has been called "egalitarian" pairing. Evidence for analogous recombination-independent pairing via multiple interstitial interactions has also been presented for Sordaria (4) and budding and fission yeast (5-12), with hints of such effects in humans (13). Additionally, in the filamentous fungi Neurospora and Ascobolus, certain repeat-recognition phenomena are proposed to involve direct pairing of the involved sequences (14, 15).

In other situations, recombination-independent pairing is found to occur at a certain specific locus. Examples include pairing between: (i) the X and Y chromosomes during male meiosis in Drosophila (3); (ii) "pairing centers" present near, but not at, the end of each chromosome during meiosis in the nematode *C.elegans* (16) and (iii) the two X chromosomes in cells of placental mammals, prior and prerequisite to inactivation of one of the two partners (17-20). In the first two cases, the involved DNA regions carry multiple non-tandem repeats; and in all three cases, specific proteins are implicated in either establishment and/or maintenance of pairing. More generally, there is a tendency for heterochromatic regions (characterized by a paucity of genes and a less "open" chromatin structure) to cluster independent of sequence homology and, in that context, to engage in sequence specific interactions that either establish or maintain a pairing contact (3, 16).

The central question for recombination-independent pairing is determining the basis for homology recognition. The most obvious possibility is direct DNA/DNA interaction between homologous sequences. Theoretical models have been proposed in which homology recognition arises from non-Watson-Crick hydrogen bond interactions

between bases in the major or minor grooves (21, 22). Local melting could also occur, permitting recognition via standard Watson-Crick base pairing. Other theories suggest that homology recognition can occur due to interactions between sequence-dependent charge distributions associated with neighboring DNA helices, where the charge distributions include not only the phosphates in the DNA but also other monovalent and/or divalent ions that are bound to or very near the neighboring DNA molecules (23-32). Interaction-induced correlations between the spatial distribution of charges can result in energy minimization when sequence-matched helices are in close proximity (24-26, 31-36, SI text). Despite its *a priori* attractiveness, acceptance of the possibility of pairing via direct DNA/DNA interactions has been impeded by the lack of conclusive experimental evidence that such a process is possible in biologically-relevant conditions. Encouragement is provided, however, by recent experiments showing evidence for preferential interactions between DNA molecules with like sequences (37, 38).

Two other general scenarios for homology recognition have also been envisioned. In one scenario, information comes from local sequence information that is read out indirectly into other determinants that mediate the actual pairing process, e.g. site-specific binding proteins which then "dimerize" in *trans* or interaction of transcription complexes and/or RNAs that then carry out the inter-chromosomal interaction (39). In the other scenario, homology is recognized along the length of a chromosome only via the spatial pattern of particular inter-chromosomal snaps which, in the most extreme case, could be identical at every position in the array (31).

The present study further investigates the possibility of direct, homology-directed protein-independent dsDNA / dsDNA pairing. We provide multiple lines of evidence for homology-dependent pairwise interactions between chemically intact DNAs. We show

that such pairing can occur under biologically sensible conditions, at positions far from the ends of the interacting molecules, in the presence or absence of non-specific competitors. Conditions known to promote nonspecific aggregation of DNA are avoided. We also begin to explore parameters of sequence length, temperature and monovalent salt concentration to provide a good basis for comparison with proposed theoretical models for pairing. This analysis was carried out with a parallel single molecule approach that uses magnetic tweezers and permits simultaneous and specific detection of hundreds of paired molecules.

The presented results should encourage further research into the nature of biologically significant cases of recombination-independent pairing and will provide an opportunity to advance the many theoretical studies that have made predictions about the functional dependence of the pairing whose differentiation and validation require experimental data.

Results

Pairing of homologous DNAs was determined using the parallel single molecule, magnetic tweezers-based, assay system described in Fig. 1 and Materials and Methods. The homology-dependent DNA/DNA pairing was investigated using a technique similar to a sandwich assay (Fig. 1). Pairing of two DNAs is monitored by measuring the number of bound beads attached to a capillary surface, where the beads are connected to the surface by the paired DNA. The two DNAs of interest are differentially labeled at "opposite" termini, one with biotin and the other with digoxigenin (Dig) (above; Fig. 2, green circles and red diamonds, respectively). The Dig-labeled DNA can attach specifically to the anti-Dig labeled capillary, and the biotin labeled DNA can attach specifically to the magnetic bead. No single DNA molecule can specifically bind a magnetic bead to the capillary surface, but if a DNA molecule labeled with Dig pairs with a DNA molecule labeled with biotin, then a paired molecule can specifically bind to both a bead and the capillary. Thus, only paired molecules can specifically bind magnetic beads to the capillary, so the number of paired molecules can be determined by counting the number of magnetic beads bound to the capillary.

The presence of homologous pairing is further documented by quantification of the distance of the tethered beads away from the surface of the slide (Fig. 1). In all experiments, pairing was examined between a Dig-labeled full-length lambda (λ) DNA and one of a variety of biotin-tailed partners (above), either full-length λ or a smaller molecule carrying a subregion of λ. If pairing detected by bead tethering is homologous, the distance of the bead to the surface should correspond to sequence matched position of the biotinylated end along the complete λ molecule. Thus, the position of beads corresponding to the pairing of two complete λ molecules is approximately the same as

the ~12 micron length of λ when extended by a force of 2 pN, whereas partners with ends that are sequence matched to regions inside λ will have shorter bead/capillary separations. In this system, paired molecules are tethered to a glass slide at one end of one partner and extended by 2 pN of force imposed on a magnetic bead tethered to the opposite end of the other partner. Pairing is detected as tethering of a bead attached to the end of one of the two molecules, with homologous pairing revealed by signals present at the appropriate diagnostic distance from the slide. Standard conditions for analysis of pairing were 150 mM NaCl phosphate buffer (PBS) at 25°C with equal volumes of the two differently labeled molecules, always at equal concentrations, which ranged from ~3-10 nM (in molecules) and 30-100 ug/ml.

A signal diagnostic of homologous λ x λ pairing emerges with time of incubation. Co-incubation of differentially-labeled λ DNAs under standard conditions results in the appearance of an appropriate diagnostic signal (Fig. 2*A*). The level of this signal initially increases with time and then plateaus. Further, at very short (10 min) incubation times, no pairing signal is observed above the non-specific background, confirming that no pairing is occurring during the tethering steps and thus that the signal observed after longer incubation times does not involve artifactual co-localization of DNAs on the beads or other features of the detection system. In most experiments, the level of DNA-DNA pairs formed as a function of incubation time approaches an asymptotic value (e.g. Fig. 2*A*). For full λ pairing with full λ the plateau probably occurs because most of the molecules in the reaction have already become paired, as discussed further below.

The λ x λ pairing signal results from pairwise association.

To test that the bead binding in the singly labeled samples is indeed due to two-by-two interactions between molecules, and not to a higher order interaction, we measured the

rate of bead tethering as a function of the DNA(s) concentration. In a pairwise interaction, the rate of formation of beads bound at the appropriate diagnostic distance should increase with the square of the (equal) concentrations of the two differentially labeled DNAs. For this analysis, we measured the number of bound beads formed after a 1 h pairing incubation. For this 1 h incubation time, the level of pairing increases linearly with time, implying that the number of bound beads corresponds to the pairing rate rather than the equilibrium value as can be seen in Fig. 2*A*. Fig. 2*B* shows that the number of beads bound after 1 h incubation increases with the square of the concentration of molecules in the pairing mixture for DNA(s) concentrations of 30 to 100 μg / ml (1- 3 nM in molecules. There is a possible slight deviation from this relationship at the highest concentration where, even at a short incubation time, the number of bound beads may have begun to saturate due to almost complete pairing in the sample (below). At concentrations below 30 μg / ml (1 nM in molecules) the number of tethered beads is of the order of the non specific binding signal. These data confirm that the observed bead-tethering signal represents a pair-wise DNA/DNA interaction.

Homologous λ x λ pairing is very efficient.

To assess the efficiency of homologous pairing we compared the number of tethered beads observed in a standard λ x λ pairing reaction experiment with that observed for a single λ with biotin attached at one end and Dig attached at the other. For this construct tethering does not require interaction with another molecule. The two samples contained the same total amounts of DNA (100 μg / ml (3 nM)), and were incubated in parallel in 150 mM NaCl at 37°C for 1 h. A typical result was 820 bound beads for the doubly labeled control DNA sample and 42 bound beads for the pairing DNA sample. Perfect pairing would correspond to a bound bead number that is 25% of that seen for the doubly

labeled DNA[*] implying that the pairing efficiency under these conditions is slightly over 20 %. An even higher level of pairing can be achieved at modestly higher DNA and salt concentrations and longer incubation times (Fig. 2*A*). Thus: homologous pairing of two 48 kb λ DNAs is an extremely efficient process, with essentially complete pairing readily achievable.

[*Footnote: (i) if all molecules were present in homologous pairs, half of all pairs would contain either two biotin-labeled molecules or two Dig-labeled molecules and would not be detected; and (ii) each tethered bead represents two λ DNAs rather than a single DNA as in the control sample.]

5 kb DNAs pair with the homologous regions of full length λ DNA.

To further explore the requirements of homologous pairing we examined the interaction between λ DNA labeled with Dig and a series of biotinylated DNAs carrying 1-to-5 kb subregions of λ corresponding to selected positions along the length of the full λ molecule.

As discussed above, an important signature of homologous pairing is the separation of the magnetic bead from the capillary, which should match the position of the λ sequence that matches to the biotinylated end of the 5 kb fragment. Histograms of the fraction of the detected beads that were located at a particular distance from the capillary are shown in Fig. 3 for three different homologous subregions of λ. In each case, pairing occurs at exactly the appropriate specific position as shown by the distribution of length for tethered beads after a 1 h pairing incubation period (Fig. 3, blue, green and purple histograms). For comparison, the distribution of bead positions is also shown for the standard λ x λ pairing reaction and for the control sample in which biotin

and Dig-labels are present at the two ends of each individual DNA (Fig. 3, yellow and gray-outlined histograms, respectively).

The 5 kb DNAs used for the above analysis comprised only the sequences of interest, leaving open the possibility that pairing might require separation of Watson-Crick base pairs at the ends of the molecule. To exclude this possibility we also examined pairing of Dig-labeled full-length λ with a molecule in which the 5 kb of λ sequence was embedded in non-homologous sequences of 400 bp and 2 kb present at bead-proximal and bead-distal ends, respectively. Pairing still occurs at the appropriate position (Fig. 3, compare magenta outline histogram with underlying blue histogram in top panel) at comparable efficiency. Pairing at the homologous position is also observed between full-length λ and homologous 5 kb DNAs with much longer tails (prepared as in Fig. S1*AB;* data in Fig. S1*C*, second row; further discussion below).

Importantly, and in contrast to the above findings, no pairing signal is observed between full-length λ and a 5 kb DNA from pcDNA3.1. As shown in Fig. S2*A*, there is no increase in the number of bound beads as a function of the incubation time, with the number of beads bound (per unit length of capillary) similar to that seen in control samples containing either of the two types of DNA alone (not shown). Further, the distribution of bead positions peaks at the length of λ (~14 μm) which is also the dominant bead position for non-specific binding of the control sample containing only the λ DNA end-labeled with Dig (Fig S2*BC*).

Pairing can be mediated by regions of homology shorter than 5 kb. Experiments analogous to those described above examined pairing between full-length λ DNA and

DNAs sharing regions of homology of 1 kb, with and without long flanking DNA tails on either side.

- In PBS, in the absence of tails, the pairing of a 1 kb λ DNA by full length λ does not result in bead binding that significantly exceeds the controls (see Fig S3).

- Molecules with long non-homologous tails flanking 2 kb and 1 kb of homology exhibit both high levels of tethered beads, comparable to those for long tailed molecules with 5 kb of homology (Fig. S1*C*, 2nd to 4th rows). Further, in all three cases, the pairing signal is absolutely dependent on the presence of homology: the number of bound beads present with a homologous regions sandwiched between long tails significantly exceeded those seen in the standard control samples as well in analogous pairing reactions containing Dig-labeled λ DNA and a biotin-labeled molecule containing only the non-homologous tail regions and no λ DNA (data not shown). Interestingly, however, all long-tailed molecules with central regions containing λ homology give broad distributions of bead positions, even when the homology region was 5kb. Further, the breadth of the distribution tends to increase as the length of the homologous region decreases, where some distributions do not even peak at the position of the homology. This spreading of the distribution appears to be related to the presence of long non-homologous tails, as it was not observed for 1 kb or 5 kb fragments lacking tails (e.g. compare first and second rows of Fig. S1C), and is not a prominent feature of a 5 kb region with shorter tails, although there may be some spreading towards shorter distances in that case (Fig. 3, top panel).

All of these data can be united by a model in which (i) formed pairs are significantly unstable; (ii) in the presence of long non-homologous tails, loss of

homology-dependent contacts leads to "sliding" of the smaller duplex, which preserves the interaction at non-homologous positions; while (iii) in the absence of adjacent non-homologous tails, loss of homologous contact results in unbinding (see also above). This model requires the existence of a short-range non-sequence specific attractive interaction which becomes important only after homologous interactions have brought the non-homologous portions into sufficient proximity for the short range non-specific interaction to become important. The existence of a short range non-sequence dependent interaction is consistent with experimental observations that have shown that dsDNA will form tightly coiled toroids; further, toroid formation can be initiated by kinks that bring parts of the dsDNA molecule close to itself (40), analogous to the proposed effect of homologous pairing in the present experiments.

Homologous pairing is increased by crowding agents and is not affected by the presence of competitor DNA or BSA. We were interested to know whether pairing would be altered by inclusion of additional factors that bring the reaction conditions closer to those present *in vivo*.

Molecular crowding. Our standard pairing reactions contain ~ 0.003 pmol of each type of DNA in a 2 μl volume, for an overall total molar concentration of homologous molecules of 3nM. This is roughly commensurate to the molar concentrations of two homologous DNAs in a single yeast cell nucleus (two molecules in nucleus of 1μm diameter and thus a volume of 10^{-15} l). It might be expected that the excluded volume effects associated with the molecular crowding *in vivo* might increase the pairing level. We therefore tested the effects of polyethylene glycol (PEG), average molecular weight 8000 daltons, which is often used as a crowding again in DNA experiments. The presence of PEG produces a 4-fold increase in the formation of DNA-

DNA pairs during short incubation times (1-5 h), implying an increase in the initial pairing rate (Fig. 4*A*). It is in good agreement with previous results pointing to homology-dependent dsDNA associations under conditions of crowding created by osmotic stress (37,38).

Non homologous competitor DNA. We find that pairing between a λ molecule and a 5 kb fragment (lacking tails) is unaffected by the presence of unrelated DNAs of several types, except at high concentrations of competitor where pairing is actually increased. (1) Pairing is unaffected by the presence of a 5 kb fragment of pcDNA3.1 at three times the concentration of the 5 kb λ fragment (in molecules and base pairs) (Fig. 4*B*). (2) Similarly, human genomic DNA containing fragments of sizes between 40 and 300 kb, with an average at 200 kb at a concentration ratio of 1:4-fold in molecules and 10:1-fold in base pairs does not affect pairing (Fig S4). (3) Finally, in the presence of fish sperm DNA, average length of about 400 bp, at a concentration ratio in molecules of 250:1 (20:1 in base pairs) relative to the 5 kb λ fragment, both the rate and final level of pairing are increased. We attribute this increase to molecular crowding at the high competitor DNA concentrations used in this experiment. We also note that the total DNA concentration (in base pairs) present in the latter experiment ~ 2 mg/ml, approaches the range DNA concentrations found in a eukaryotic nucleus *in vivo* (10-50 mg/ml). Thus: homologous pairing is not decreased by the presence of competitive non-homologous dsDNA at concentrations substantially exceeding those of the homologous dsDNA, for competitors having lengths longer, shorter, or equal to the length of the 5 kb homologous DNA partner.

Protein. We probed the effect of proteins on the pairing reaction by including 0.1

% bovine serum albumin (~15μM). Again, no evident effect on pairing is observed (Fig.

S4).

Taken together these results show that non homologous sequences do not compete

for the homologous sites, that homologous pairing is not suppressed in complex

environments; and that molecular crowding increases the rate and efficiency of pairing.

Thus, although we cannot fully reproduce the complex cellular environment found in

vivo, these observations suggest that homology-dependent DNA/DNA pairing is not

impeded by the presence of non-homologous DNAs and will be favored by the crowded

conditions characteristic of the cell nucleus.

Comparison with Watson-Crick pairing. The observations presented above demonstrate

that pairing can occur in regions far from the ends of either interacting molecule, and

therefore imply that (a) pairing is not attributable to melting/fraying that gives rise to

open ends and thus (b) is not attributable to simple Watson-Crick reannealing at such

ends. We have further probed the possible involvement of ssDNA in the observed pairing

reaction in four ways.

First, we asked whether λ x λ pairing is affected by the presence of T4 gene32

ssDNA binding protein, which can bind to as few as 8 nt of ssDNA. If ssDNA were to

play a significant role, occlusion by gene 32 protein should inhibit the reaction. We

observe, however, that pairing is the same in the presence and absence of this molecule

(data not shown), implying that ssDNA interactions are not involved in the initial

formation of pairs.

Second, we measured the ability of base pairing interactions to mediate pairing.

We examined the formation of dimers between two λ phage DNAs, one labeled with

biotin and the other with Dig at its "opposite" end, and carrying complementary 12 bp ssDNA overhangs at their respective unlabeled ends. After a 2 h incubation under standard pairing conditions (60 μg / ml (~ 2 nM) of each DNA) the number of dimers was only 20% of the number of DNA-DNA pairs measured in parallel in our standard λ x λ reaction under the same conditions, again suggesting Watson-Crick binding does not play a significant role.

Third, to test whether pairing might involve Watson-Crick binding within bubbles of ssDNA along the dsDNA, we heated a λ x λ pairing mixture to 50°C, and then quickly quenched it, prior to incubation for pairing. This procedure should create ssDNA bubbles, preferentially in AT-rich regions, that could pair with other open bubbles in such regions. However, no increase in the rate or level of pairing was observed, further arguing that ssDNA bubbles are unlikely to be involved.

Fourth, regions of Watson-Crick duplex even as short as 12 bp are resistant to shear forces up to 25 pN (41). In contrast, molecules linked by homologous pairing over regions of 5 kb or 48.5 kb shear at forces of 10-20 pN. Thus, homologous associations are not maintained by single or multiple Watson-Crick base-pairing interactions of 12 bp or longer, although involvement of even shorter regions of base-pairing is not excluded.

We also note that the involved pairing likely does not involve interactions between matching sites of deformed structure along the two DNAs. The long-tailed DNA segments that pair efficiently with full-length λ DNA were generated from *in vivo* amplified plasmid DNA, and terminally biotinylated subfragments were prepared without the use of ethanol or any chemical (e.g. phenol) that is known to generate such deformations (Fig. S1*A*).

Pairing as a function of temperature and monovalent and divalent salts. We have also used the λ x λ pairing reaction to begin to probe the effects of temperature and of salt on the observed pairing process.

Temperature. The level of pairing increases significantly as the temperature is raised from 5ºC to 40ºC, and then decreases strongly at higher temperatures, with the same relationship observed both before and after the reaction has plateaued (Fig. 5*A*). Preliminary experiments further suggest that the width of the bead location distribution does not increase for temperatures between 40ºC and 60ºC, even though the number of paired molecules decreases dramatically. Thus the decreased signal at higher temperatures still corresponds to regular homologous pairing rather than replacement of a homologous interaction by a non-homologous one. More generally, higher temperature does not promote non-homologous pairing.

Monovalent and divalent salt. The rate of pairing, defined by pairing levels after a 1 h incubation, increases monotonically with the level of either NaCl or KCl over a range of concentrations from 50 mM to 1 M, with slight differences for Na^+ versus K^+ (Fig. 5*B*). Below 50 mM (15 mM), pairing is of the order of the non-specific signal. Divalent salts are well-known to promote aggregation of DNA; however, it has not been clear whether such aggregation could be homology-dependent or not (42). We therefore also examined homologous pairing in reaction mixtures where $MgCl_2$ is present instead of a monovalent salt. We find significant pairing in 10 mM $MgCl_2$ during a 10 min incubation time, whereas pairing is almost absent in 10 mM NaCl even after an hour of incubation. Finally, no pairing was observed when pcDNA3.1 was paired with full λ in 10 mM $MgCl_2$, further demonstrating that even in $MgCl_2$ the pairing is sequence dependent.

Implications. These findings are compatible with occurrences of homologous DNA/DNA pairing *in vivo*. Pairing occurs robustly throughout the range of temperatures encountered by most living organisms and at salt concentrations corresponding to those generally thought to occur *in vivo*: 150 mM monovalent salt, with K^+ predominant over Na^+; and ~10 mM total Mg^{2+}, with the concentration of free Mg^{2+} being considerably lower (43). The patterns of effects of temperature and salt are also of interest in relation to physical mechanisms of pairing (below).

Discussion

The results presented reveal that two homologous DNA segments can efficiently and rapidly identify one another and interact to form complexes stable against thermal motion, in the absence of proteins. These findings confirm and extend the results of two other studies pointing to such a possibility (37, 38). Both of these experimental studies include extensive discussions of theories that predict sequence dependent pairing. A detailed review of those proposals is not possible in this work, but we recommend them as an excellent source of information on possible mechanisms for sequence dependent pairing. The most important conclusion from the current work is that direct DNA/DNA interactions occur under physiologically sensible conditions; therefore such DNA/DNA interactions may underlie recombination-independent pairing *in vivo*.

The Mechanism of Homology-dependent dsDNA/dsDNA Pairing.

General Aspects. Any process that leads to persistent homology-dependent association of two DNA duplexes will be characterized by certain features. A further discussion of theoretical studies can be found in SI text.

First, bringing two dsDNA molecules together in solution must include one or more attractive interactions that overcome strong inter-molecular repulsion due to the negative charges on the phosphodiester backbones; therefore, the effectiveness of attractive interactions will be increased if the repulsion is reduced. Increased salt concentrations would be expected to reduce the repulsion, but of course, changing the salt concentration may alter other interactions as well so further theoretical work is required in order to predict the salt dependence of the attractive interaction. We show that the pairing rate increases with salt and that at low concentrations of salt, the pairing rate in Na^+ significantly exceeds the pairing rate in K^+. We note that earlier work that showed that at low concentrations Na+ screens the intra-molecule backbone repulsion better than K^+ (44). Similarly, Mg^{2+} at 10 mM concentrations promotes pairing more effectively than either monovalent cation at 150 mM. We note that earlier work has shown that that melting temperature and unzipping force in 10 mM Mg^{2+} are comparable to those in 150 mM NaCl, where the melting temperature and unzipping force also depend on the intra molecular backbone repulsion. (45,46) These findings show that the pairing is not simply dependent on the ionic strength of the solution, and suggest that the pairing may be correlated with the measured values that are related to the screening of the intra molecular backbone repulsion.

Second, the attractive interaction that brings pairs together must be dominated by homology-dependent forces, since we do not observe pairing between sequences without homology and the pairing of homologous DNA is not suppressed by the presence of non-

homologous competitors with lengths that are longer, shorter, or equal to that of the homolog. Having the long range attractive interaction dominated by homology is important in avoiding unwanted non-homologous interactions.

Third, interactions between homologous regions must be must be strong enough to allow correctly bound sequences to remain together but weak enough that unmatched sequences with small regions of accidental homology unbind rapidly, thereby avoiding kinetic trapping in non-homologous interactions.

Specific mechanisms for homology recognition. Two types of recognition mechanisms have been proposed: direct mechanisms in which attractive interactions involve the bases themselves (21) and indirect mechanisms where recognition and attraction occur through interactions involving the helical structures of the molecules that vary in correlation with base pair sequence (24-26, 31, 32, 34).

While extensive further theoretical and experimental work is required before firm conclusions can be formed about the processes underlying the pairing demonstrated in this work, one observation does help to discriminate among models proposed thus far. We observe that the rate and extent of pairing increases progressively with temperature up to 40°C and then decreases. In general, indirect pairing models depend on the matching of the conformations of two neighboring dsDNA molecules, and it has been predicted that conformation-based models should exhibit exactly the pattern of temperature dependence that we observe. At low temperatures, homologous molecules may get frozen in non-matching conformations that do not result in a strong attraction; however, higher temperatures may allow molecules to readjust conformations in response to their neighbors, permitting the two molecules to minimize their energy by pairing. (31). At still higher temperatures, the conformational fluctuations may be too large to

allow the molecules to adjust and pair. We note that the sequence specificity of the pairing does not decrease with temperature. We also note that hydration has been proposed as a possible pairing mechanism, where homologous pairing results in the displacement of specifically bound water into the disordered solvent in an entropy-driven process (31). Such sequence dependent hydration effects are already known to play a role in sequence dependent DNA-protein interactions (47) and RNA folding (48), so it would be reasonable to assume that hydration effects also play a role sequence dependent dsDNA/dsDNA interaction.

Implications for recombination-independent pairing <u>in vivo.</u> The conditions under which pairing is observed in the present study are fully compatible with occurrence of analogous interactions *in vivo*. Pairing occurs efficiently between relatively short regions of homology; independent of DNA ends, under physiological conditions of monovalent salt; in the presence of complex nonhomologous competitor DNA and nonspecific protein; and over reasonable time scales at concentrations of the two DNA segments that are comparable to the concentrations for two homologous segments in a yeast nucleus, with higher rates (or comparable rates at lower DNA concentrations) achievable by inclusion of crowding agents and by optimization of temperature.

Further, the DNA concentrations involved in the observed pairing are orders of magnitude lower than those required for collapse of DNA into toroids in NaCl via nonspecific attractive interactions (40). Additionally, pairing is observed for full-length λ DNAs and for shorter DNAs prepared in such a way as to avoid potential artifacts that could result from use of PCR-generated fragments.

Taken together the presented findings suggest that direct homologous DNA/DNA interactions could be responsible for guiding recombination-independent pairing of homologous DNAs *in vivo*.

Furthermore, the robustness of the observed pairing process raises the strong suspicion that intrinsic homology-based pairing interactions may be the "default option" *in vivo*. By this view, evolution has specifically ensured that pairing between homologous chromosomes is precluded genome-wide, with restrictions then lifted specifically in specialized cases where pairing is useful or, in a few organisms, left in place and accommodated as somatic pairing. In this scenario, specific genetically-encoded "pairing determinants" would, at least in part, be features that counteract negative, inhibitory constraints present more typically through the genome. This notion is diametrically opposed to the common view that homologous pairing is a rare and unfavorable condition that must be specifically promoted by appropriate biological features. However, specific proteins could be required to stably maintain DNA/DNA pairing once it has occurred (below).

The possibility of robust DNA/DNA interactions is also interesting in the context of the overall process of homology searching. It has been argued that the most significant problem for such a process is not the need to find the correct partner but rather the need to avoid getting kinetically trapped in stable non-specific interactions, which in turn requires any general homology searching process must involve weak, transient "kissing" interactions (49). Recent theoretical work and modeling on macroscopic systems supports the usefulness of a pairing process that occurs in stages, where the first stage requires weak transient interactions to avoid trapping in incorrect pairs. The initial weak binding of short matching regions should be strong enough to allow neighboring regions to bind if

they are matched, but weak enough that the two short matching regions will unbind if the neighboring regions do not match. Under these conditions, correctly aligned homologs will rapidly form strong bonds, but short regions of accidental homology will not trap pairs in false minima. Additional recognition stages can promote sequence stringency over longer and longer lengths (50). Modulation of the strength of DNA/DNA interactions could conveniently achieve such an outcome. In accord with this possibility, in two of the most robust cases of "site-specific" pairing, for the XY chromosomes of Drosophila and for "pairing centers" in *C. elegans*, specificity is conferred by clusters containing multiple non-tandem repeats of short sequences (3,16). Further, for Drosophila, where 50 copies of a <250bp repeat are normally involved in pairing, it has been shown that eight copies are largely sufficient to confer detectable pairing but two copies are not (3).

Conversely, sticky sequence-specific protein factors would seem unsuitable as primary mediators of homology recognition and thus might be envisioned either as factors that enhance the susceptibility of underlying sequences to DNA/DNA interaction and/or as factors that further stabilize contacts made at the DNA level. Correspondingly, for Drosophila somatic pairing, on the scale of an entire chromosome, an effective pairing contact, once formed, is quite stable (51).

It is also interesting to consider the possibility of direct DNA/DNA pairing interactions as a factor in interactions between sister chromatids, acting prior to or in concert with the known factors of protein-mediated cohesion and topological linkages. Sister chromatids will automatically tend to emerge from a replication complex into a confined joint space, and at least transiently, lack a full complement of nucleosomes, features that could favor DNA/DNA interactions between sisters.

Conclusion

The current observations show that homologous dsDNAs can specifically recognize one another and pair stably enough for detection on a time scale of minutes/hours. These findings encourage future studies to assess the physical basis for such homology-dependent recognition and the relevance to, and rules for, DNA/DNA-mediate homlogous pairing *in vivo*.

Materials and Methods

    Sample preparation. Lambda phage DNA (NEB, Beverly, MA) (48502 base pairs) as well as DNA molecules that contain subregions of λ DNA with or without dsDNA tails of non-λ sequences were used in this study. To assay pairing between two full-length λ DNAs, two types of samples were prepared. In one case, λ DNA was hybridized and ligated to an oligonucleotide complementary to the ssDNA tail at the left end of λ that contained a Dig-label (5'Dig TTT TCC AGC GGC GGG 3') and to another oligonucleotide without label at the right end (5' AGG TCG CCG CCC 3'). In the other case, λ DNA was hybridized and ligated to a biotinylated oligonucleotide (AGG TCG CCG CCC TTT Biotin 3') at right end and to an oligonucleotide without label at the left end (5' TCC AGC GGC GGG 3'). For assessing pairing of λ with specific subregions, smaller DNA molecules containing the subregion of interest carried a terminal biotin label at their "right-proximal" end interacted with the Dig-labeled full-length λ DNA described above. To examine pairing with DNAs comprising entirely homologous DNA, three different 5 kb fragments, corresponding to different segments of the λ genome, were prepared using high fidelity PCR by amplification starting at positions 116, 11302, and 21613. To examine DNAs containing one of these 5 kb subregions, but carrying (long or short) tails of non-λ sequences, a different preparation procedure was used which, concomitantly, minimizes the possible introduction of aberrancies in the dsDNA duplex structure (details in SI Materials and Methods and Fig. S1).

    Assay.

    This assay system has the potential to detect up to ~1000 tethered beads in parallel in a single assay (52). Under standard conditions, the initial imposed force is 2 pN. As this may be sufficient to dissociate some weakly bound pairs, the assay should be

viewed as a measure of the number of pairs that are bound sufficiently strongly to withstand 2 pN of shear force.

Pairing reaction protocol. Initially equal volumes of each sample are mixed and incubated for a chosen period of time, varying from 10 min to 7 days, where most experiments had incubation times of a few hours. At the end of the incubation time, an aliquot of the incubation mixture is then incubated for 2 min at 37ºC with superparamagnetic (Dynal 2.8 μm diameter) streptavidin-coated magnetic beads. Subsequently this solution is placed for 10 min in a microchannel with square cross-section 0.8 mm, containing a round capillary, 0.55 mm diameter, previously coated with anti-Dig antibody (53) During this last step, Dig-labeled molecules become tethered to the surface of the capillary while biotinylated molecules remain associated with magnetic beads. Importantly, the 12 min tethering step is much shorter than the incubation time that is required for a pairing signal to appear (1 – 24 h; below); consequently, pairing during tethering makes little or no contribution to the total level of observed pairing. Finally, a force is applied to the beads by bringing a permanent magnet close to the capillary to apply a constant force of 2 pN. Free beads move toward the magnet and are lost from the field of view (Fig. 1*A*) Beads that are bound by the pairing between one molecule with a Dig-labeled end and another molecule with a biotin labeled end are detected (Figs. 1*AB*), and the number of tethered beads separated from the edge of the inner capillary by a DNA-DNA pair is counted.

To maximize reproducibility in our measurements, the mixture of labeled dsDNA samples is separated in several tubes prior to incubation and every sample assayed for pairing level is taken as an aliquot from a different tube. This procedure avoids variations due to shear forces created by pipetting from the same tube several times.

Further, to minimize variation in the efficiencies of association of terminal tags with the bead and/or the antibody-coated slide, comparisons of pairing efficiencies between/among different situations were always carried out in parallel, on the same day, with the same preparations of all non-DNA reagents.

References

1. Kowalczykowski SC (2008) Structural biology: snapshots of DNA repair. *Nature* 453:463-466.

2. Coic E, Feldman T, Landman AS, Haber JE (2008) Mechanisms of Rad52-Independent Spontaneous and UV-Induced Mitotic Recombination in *Saccharomyces cerevisiae*. *Genetics* 179:199-211.

3. McKee BD (2009) in in *Meiosis*, eds Benavente R, Volff J-N (Karger Press, Basel), pp 56-68.

4. Storlazzi A et al. (2003) Meiotic double-strand breaks at the interface of chromosome movement, chromosome remodeling, and reductional division. *Genes Dev* 17:2675-2687.

5. Weiner BM, Kleckner N (1994) Chromosome pairing via multiple interstitial interactions before and during meiosis in yeast. *Cell* 77:977-991.

6. Cha RS, Weiner BM, Keeney S, Dekker J, Kleckner N (2000) Progression of meiotic DNA replication is regulated by interchromosomal interaction proteins, negatively by Spo11p and positively by Rec8p. *Genes Dev* 14:493-503.

7. Keeney S, Kleckner N (1996) Communication between homologous chromosomes: genetic alterations at a nuclease-hypersensitive site can alter mitotic chromatin structure at that site both in *cis* and in *trans*. *Genes Cells* 1:475-489.

8. Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295:1306-1311.

9. Burgess SM, Kleckner N, Weiner BM (1999) Somatic pairing of homologs in budding yeast: existence and modulation. *Genes Dev* 13:1627-1641.

10. Nabeshima K, Kakihara Y, Hiraoka Y, Nojima H (2001) A novel meiosis-specific protein of fission yeast, Meu13p, promotes homologous pairing independently of homologous recombination. *EMBO J* 20:3871-3881.

11. Scherthan H, Bähler J, Kohli J (1994). Dynamics of chromosome organization and pairing during meiotic prophase in fission yeast. *J Cell Biol* 127:273-85.

12. Molnar M, Kleckner N (2008) Examination of interchromosomal interactions in vegetatively growing diploid *Schizosaccharomyces pombe* cells by Cre/loxP site-specific recombination. *Genetics* 178:99-112.

13. LaSalle JM, Lalande M (1996) Homologous association of oppositely imprinted

chromosomal domains. *Science* 272:725–728.

14. Selker EU (1999) Gene silencing: repeats that count. *Cell* 97:157-160.

15. Rossignol JL, Faugeron G (1994) Gene inactivation triggered by recognition between DNA repeats. *Experientia* 50:307-317.

16. Bhalla N, Dernburg AF (2008) Prelude to a division. *Annu Rev Cell Dev Biol* 24:397-424.

17. Augui S et al. (2007) Sensing X chromosome pairs before X inactivation via a novel X-pairing region of the Xic. *Science* 318:1632-1636.

18. Xu N, Tsai CL, Lee JT (2006) Transient homologous chromosome pairing marks the onset of X inactivation. *Science* 311:1149-1152.

19. Xu N, Donohoe ME, Silva SS, Lee JT (2007) Evidence that homologous X-chromosome pairing requires transcription and Ctcf protein. *Nat Genet* 39:1390-1396.

20. Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT (2009) The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* 460:129-132.

21. McGavin S (1971) Models of specifically paired like (homologous) nucleic acid structures. *J Mol Biol* 55:293-298.

22. Wilson JH (1979) Nick-free formation of reciprocal heteroduplexes: A simple solution to the topological problem. *Proc Natl Acad Sci USA* 76:3641-3645.

23. Ramos JE, deVries R Jr., Neto JR (2005) DNA Ψ-condensation and reentrant decondensation: effect of the PEG degree of polymerization *J Phys Chem B*, 109: 23661–23665.

24. Kornyshev AA, Leikin S (2001) Sequence Recognition in the Pairing of DNA Duplexes. *Phys Rev Lett* 86:3666-3669.

25. Kornyshev AA, Leikin S (1999) Electrostatic zipper motif for DNA aggregation. *Phys Rev Lett* 82:4138-4141.

26. Kornyshev AA, Lee DJ, Leikin S, Wynveen A (2007) Structure and interactions of biological helices. *Rev Mod Phys* 79:943-996

27. Gelbart, WM, Bruinsma RF, Pincus PA, Parsegian VA (2000) DNA-inspired electrostatics. *Physics Today* 53: 38-44

28. Howerton SB, Sines CC, vanDerveer D, Williams LD (2001) Locating monovalent cations in the Grooves of B-DNA. *Biochemistry* 40:10023-10031

29. Chiu TK, Dickerson RE (2000) One angstrom crystal structures of B-DNA reveal sequence specific binding and groove-specific bending of DNA by magnesium and calcium. *J Mol Biol* 301: 915-945

30. Hud, N (2008) Nucleic Acid Metal Ion Interactions (The Royal Society of Chemistry, Cambridge, UK).

31. Leikin S, Parsegian VA (1994) Temperature-induced complementarity as a mechanism for biomolecular assembly. *Proteins* 19:73-76.

32. Kornyshev AA, Wynveen A (2009) The homology recognition well as an innate property of DNA structure. *Proc Natl Acad Sci USA* 106:4683-4688.

33. Grønbech-Jensen N, Mashl RJ, Bruinsma RF, Gelbart WM (1997) Counterion-induced attraction between rigid polyelectrolytes. *Phys Rev Lett* 78:2477-2480.

34. Wynveen A, Lee DJ, Kornyshev AA, Leikin S (2008) Helical coherence of DNA in crystals and solution. *Nucleic Acids Res* 36:5540-5551.

35. Cherstvy AG, Kornyshev AA, Leikin S (2004) Torsional deformation of double Helix in interaction and aggregation of DNA. *J Phys Chem B* 108**:**6508-6518.

36. Kornyshev AA, Wynveen A (2004) Nonlinear effects in torsional adjustment of interacting DNA. *Phys Rev E Stat Nonlin Soft Matter Phys*, 69:041905.

37. Baldwin GS, et al. (2008) DNA double helices recognize mutual sequence homology in a protein free environment *J Phys Chem* 112: 1060-1064.

38. Inoue S, Sugiyama S, Travers AA, Ohyama T (2007) Self-assembly of double-stranded DNA molecules at nanomolar concentrations *Biochemistry* 46: 164-171.

39. Xu M, Cook PR (2008) The role of specialized transcription factories in chromosome pairing. *Biochimica et Biophysica Acta* 1783:2155-2160.

40. Hud, NV, Vilfan ID (2005) Toroidal DNA condensates: Unraveling the fine structure and the role of nucleation in determining size. *Annu Rev Biophys Biomol Struct* 34,:295-318.

41. Hatch K, Danilowicz C, Coljee V, Prentiss M (2008) Demonstration that the shear force required to separate short double-stranded DNA does not increase significantly with sequence length for sequences longer than 25 base pairs. *Phys Rev E* 78:011920 1-4.

42. Angelini TE, Liang H, Wriggers W, Wong GCL (2003) Like-charge attraction between polyelectrolytes induced by counterion charge density waves. *Proc Natl Acad Sci USA* 100: 8634-8637.

43. Terasaki M, Rubin H (1985) Evidence that intracellular magnesium is present in cells at a regulatory concentration for protein synthesis. *Proc Natl Acad Sci USA* 82:7324-7326.

44. Vlassakis J et al. (2008) Probing the mechanical stability of DNA in the presence of monovalent cations. *J Am Chem Soc* 130:5004-5005.

45. Lee CH, Danilowicz C, Coljee VW, Prentiss M (2006) Impact of magnesium ions on the unzipping of λ-phage DNA. *J Phys Condens Matter* 18: S205-S213.

46. Nakano S, Fujimoto M, Hara H, Sugimoto N (1999) Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Research*. 27: 2957-65

47. Jayaram B, Jain T (2004) The role of water in protein-DNA recognition. *Ann Rev Phys Biomol Struc* 33:343-361.

48. Roh JH et al. (2009) Dynamics of tRNA at different levels of hydration. *Biophys J* 96: 2755-2762.

49. Kleckner N, Weiner BM (1993) Potential advantages of unstable interactions for pairing of chromosomes in meiotic, somatic, and premeiotic cells. *Cold Spring Harb Symp Quant Biol* 58:553-565.

50. Feinstein E. (2009) Self-assembly from the nanoscale to the mesoscale: Applications to 1D, 2D, and 3D tissue engineering, and to DNA sequence recognition. PhD Thesis, Harvard University, Cambridge MA.

51. Fung JC, Marshall WF, Dernburg A, Agard DA, Sedat JW (1998) Homologous chromosome pairing in *Drosophila melanogaster* proceeds through multiple independent initiations. *J Cell Biol* 141:5-20.

52. Danilowicz C et al. (2003) DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc Natl Acad Sci USA* 100:1694-1699.

53. Danilowicz, C et al. The structure of DNA overstretched from the 5'5' ends differs from the structure of DNA overstretched from the 3'3' ends. Accepted in *Proc Natl Acad Sci USA*

Figure Legends

Fig. 1. Pairing of homologous DNAs using the parallel single molecule, magnetic tweezers-based assay. (A) Experimental approach. The black lines represent dsDNA molecules and the two samples are distinguished by red diamonds or green circles attached to their ends corresponding to Dig- and biotin labels, respectively. The dsDNA samples after incubation are mixed with superparamagnetic beads (gray circles) and incubated inside a capillary. (B) Image showing a black region that corresponds to the capillary. The beads separate from the inner capillary at a distance of about 15 μm. The asterisks show beads that are out of focus because their corresponding molecules are tethered to positions other than the edge of the (round) capillary that is in focus. Out of focus beads are not counted in the assay.

Fig. 2. Pairing of Dig-labeled λ DNA and biotin labeled λ DNA. (A). Number of tethered beads vs. time at 2 pN, 87 μg / ml λ DNA in PBS and incubated at 37ºC. (B). Number of tethered beads vs. square of the DNA concentration, at 2 pN.

Fig. 3 Effect of sequence on DNA pairing. Distribution of extensions for about 100 beads at 2 pN for Dig-labeled λ phage dsDNA incubated with biotinylated molecules subregions from λ phage: 5 kb fragment bp # 21613 (blue) and 5 kb fragment presenting tails (magenta outline); 5 kb fragment bp # 11302 (green); 5 kb fragment bp # 116 (purple); biotinylated λ phage (yellow). λ phage molecules with both labels are shown for comparison (gray outline). The schematic representations are shown to the right of each panel.

Fig. 4 Intermolecular pairing between 6 μg/ml biotinylated 5 kb fragments and 60 μg/ml Dig-labeled λ phage in the presence of crowding agent and non-specific competitors. (A) Pairing of Dig-labeled λ DNA with 5 kb subregion bp # 21613 with and without 15 % p/v PEG 8000. (B) Pairing of Dig-labeled λ DNA with 5 kb subregion bp # 21613 in the presence of high concentration of fish sperm DNA (orange), pcDNA3.1 (purple) and control without competitor (blue).

Fig. 5 Effect of temperature and salt on the DNA-DNA interaction. (A). Number of tethered beads vs. temperature for biotinylated λ phage and Dig-labeled λ phage, 83 µg / ml in PBS incubated during several time intervals at 37ºC. (B). Number of tethered beads vs. salt concentration for 60 µg / ml biotinylated λ phage and 60 µg / ml Dig-labeled λ phage incubated for 1 hour at 37ºC in phosphate buffer 10 mM: NaCl (blue) and KCl (red).
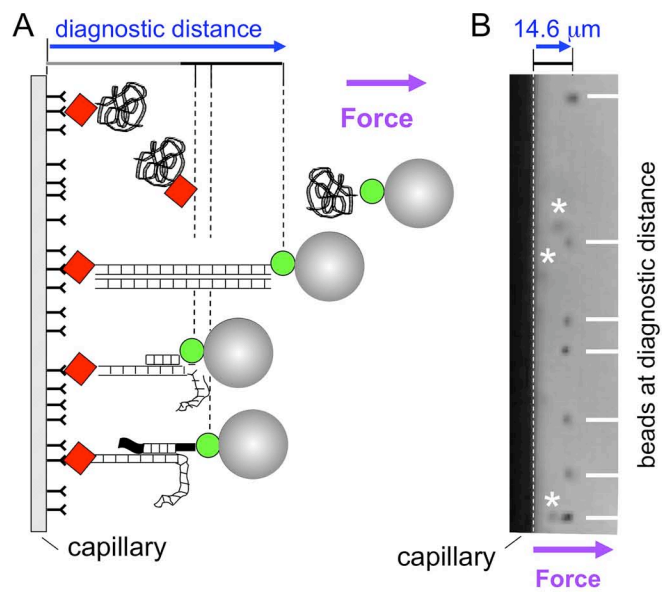
A **diagnostic distance**

Force

capillary

B 14.6 μm

beads at diagnostic distance

capillary

Force
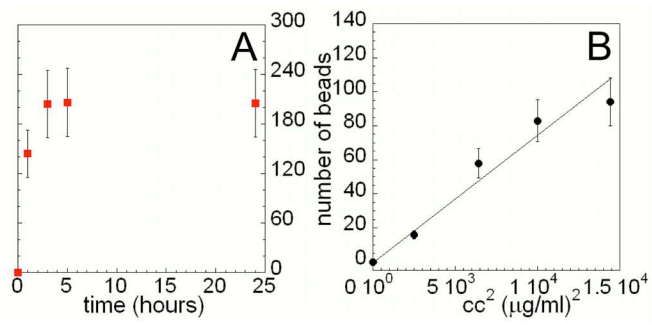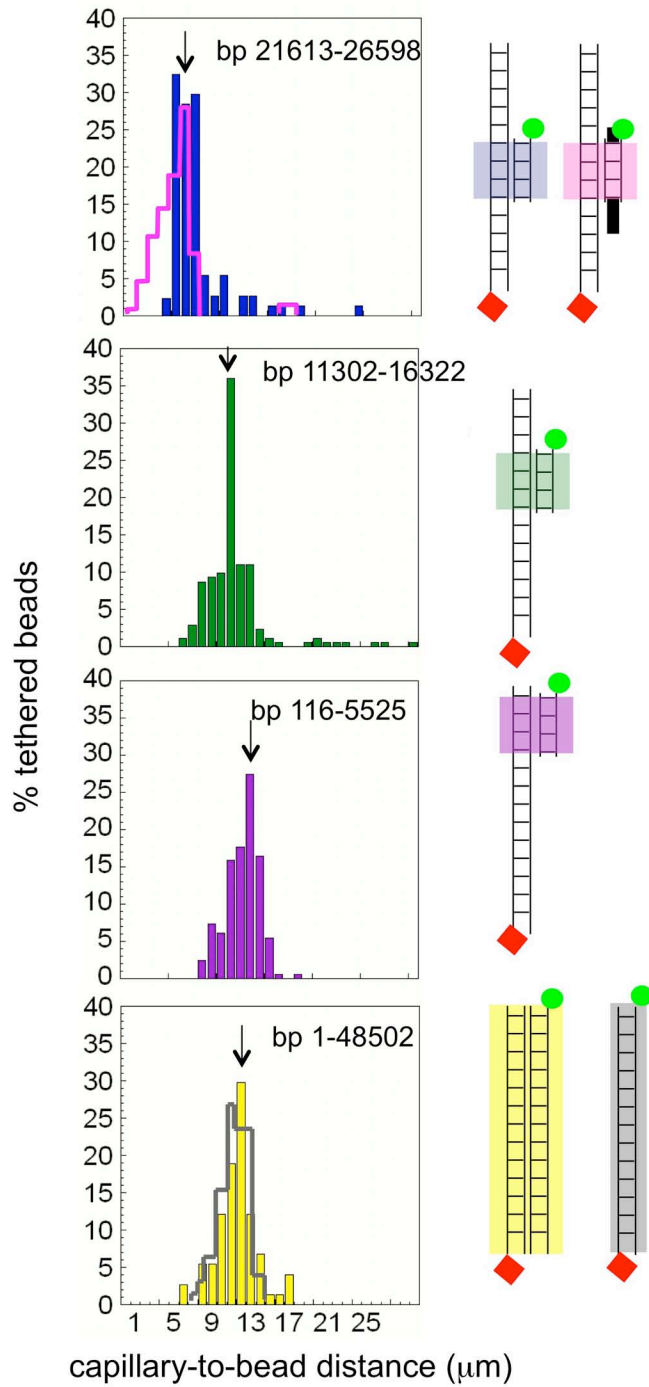
Figure 1

Figure 2

Figure 3
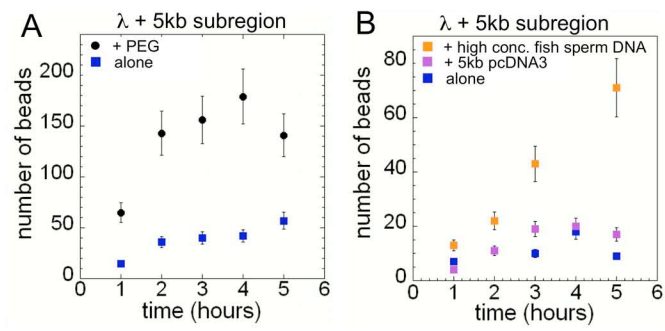
Figure 4
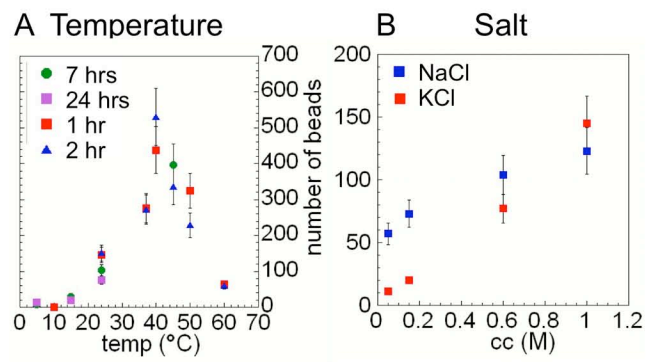
Figure 5

Supporting Information

Review of Theoretical Studies of dsDNA/dsDNA Interactions

To discuss possible mechanisms of DNA/DNA pairing, it is first necessary to consider (i) the attractive interactions that brings two DNA segments together in solution given that these molecules are strongly negatively charged and so would be expected to repel each other; (ii) whether the attractive interaction is present for any two molecules or only two molecules with sequence homology; (iii) whether additional interactions occur once the molecules are brought sufficiently close together; (iv) whether the additional interactions are sequence dependent. Finally, to interpret the results of our assay it is important to understand that we only measure the pairing between molecules that are bound sufficiently tightly to remain stable under 2 pN of shear stress, so the assay may miss paired molecules that were more weakly bound.

In a vacuum, two identical objects with fixed homogeneous charge distributions will always repel each other due to electrostatics; however, if the objects have a spatially varying charge distribution that includes regions of positive charge separated by regions of negative charge, the objects can attract each other when the separation between them is of the order of the characteristic size of the spatial variation in the charge distribution, though at larger separations the objects will still repel each other. The spatial variation can be static or dynamic. The attractive interaction between identical salt crystals is a case where the charge distributions are static. For such crystals, the attractive force is only significant when the spacing between the crystals is of the order of the lattice spacing, and the attractive force depends strongly on the exact alignment of the two crystal lattices to form one single larger crystal. Misaligned or mismatched lattices will

not result in an attractive force. The van der Waals attractive force associated with London dispersion is a case where the spatial variation in the charge distribution is dynamic and due to correlated fluctuations in the charge distributions of neighboring objects.

If two identical objects are immersed in a liquid containing counterions the interactions are much richer. Hydration forces (1) and the pressure exerted by crowding agents (2,3) can provide significant effective attractive forces that do not occur for objects in a vacuum.   In addition, even if the objects have a fixed homogeneous charge distribution, the charge distribution of the mobile counterions in solution can create a spatially dependent charge distribution on or around the object (4-6), so that the net force can be either attractive of repulsive (7). The spontaneous organization of point-like, multivalent counterions into and orderly Wigner-crystal-like lattice can provide an attraction. (7,8). A recent review has suggested that such models are not relevant for attractions between DNA molecules in monovalent or divalent ions in water at room temperature (4). Double stranded DNA in solution is a very highly charged helical molecule with a linear charge density of approximately -0.6 e/A or equivalently one negative charge every 0.17 nm where the charge is distributed along the two backbones of a helix containing a major and a minor groove with widths of approximately 2 nm and 1.4 nm, respectively; therefore, dsDNA will have a spatially varying charge distribution even without Wigner crystallization. It is also known that the exact helical structure of dsDNA (9,10) and the location of counterions in the helix can depend on sequence (11-13) Thus, homologous sequences would have matching charge distributions, but mismatched sequences would require significant DNA deformation to match the charge

distributions (4-6); consequently, the sequence dependent structure of dsDNA may provide an attractive electrostatic interaction that will be much stronger when sequences are matched. Fluctuations in the spatial distribution of bound counterions and DNA backbone may result in both sequence-dependent and sequence independent attractive interactions that initially increase with temperature and then subsequently decrease with temperature, but the attraction between homologous ones should be stronger since it does not require deformation to match the charge distribution. The low temperature increase is due to an increase in the variation in the spatial distributions of charges that allows the two structures to readjust in response to each other, and the high temperature decrease is due to excessive deviations between the spatial structures of the two molecules. (4,14,15).

In what follows, we will review the results of some previous experiments that have observed attractive interactions between dsDNA molecules. Many experiments have shown that despite the large average negative charge on dsDNA, the spatial variations in the charge density that result from interactions with multivalent and polyvalent couterions, can condense a dsDNA molecule so it collapses on itself, often forming an orderly toroid; therefore, there must be a net attractive interaction between different parts of the same highly charged dsDNA molecule (16,17). In divalent salts, there is an attractive force where some divalent ions such as Mn^{2+} and Cd^{2+} can condense individual dsDNA molecules, whereas others such as Ca^{2+} and Mg^{2+} do not result in condensation; however, divalent ions that do not result in condensation can still result in aggregation (16). Some studies that appear to consider ions with a higher valency may be misleading: for example, spermine has 4 charged groups with one positive charge each, but it is not a 4+ point-like ion. At low concentrations multivalent counterions totally neutralize the

surface charge whereas at larger concentrations they invert the sign the net macroion charge (18).

Previous experiments have not observed the aggregation of dsDNA molecules in monovalent salts even at 1 M salt concentrations unless crowding agents, such as PEG, were present to provide an attractive interaction (19,20); however, though some theory has predicted that an attractive interaction could occur in monovalent salts in the absence of crowding agents if temperature is sufficiently low, the temperatures involved so far were below the freezing point of water so the results are not relevant to in vivo conditions or in vitro experiments done in liquid water (21).

Of course sequence dependent pairing requires more than an attractive interaction. It requires that there be a sequence specific recognition step, which may or may not be incorporated in the initial attractive interaction that is required to overcome the Coulomb repulsion that is present at long distances. The possibility that the sequence recognition is included in the initial long range attractive interaction is supported by calculations (15,22) that show that the sequence dependence of the structure of the DNA backbone results in the charge distribution forming a sequence dependent bar code, where Coulombic forces and hydration forces may play a role in a sequence dependent attractive interaction (4,6,15,22). Counterion binding to DNA is also sequence-specific, potentially enhancing this barcode (13). In contrast, it is possible that an initial weak long range attraction is not sequence dependent, but that there is a subsequent step in which there is an interaction between the bases that is stable if and only if the bases match. Possible mechanisms include both Watson and Crick pairing and non-Watson Crick pairing (23-25). Thus, for the model in which the initial attraction is sequence

independent but the second interaction is sequence dependent it is assumed that the initial weak interaction will not hold molecules together in the absence of the second sequence dependent interaction.

Simple theories that do not consider water and assume that pairing depends on the static matching of Wigner crystalline structures, predict that the pairing should either increase or decrease monotonically with temperature as the ratio of the Coulombic potential to the thermal energy either increases or decreases with temperature (7); however, the interactions with water make predictions of the temperature dependence of the pairing more challenging, so more theoretical work would be required to provide a prediction of the temperature dependence of a Wigner-crystal based interaction in the presence of water. In contrast, theories that suggest that the sequence dependent pairing is a function of correlations between the dynamic charge distributions of the two helices do make clear predictions about the temperature dependence of pairing. These theories predict that the pairing will show an initial increase with temperature followed by a decrease with temperature at higher temperatures (4). Our experimental results suggest that correlations between dynamic charge distributions do enhance pairing.

Supporting Materials and Methods

Apparatus. In our apparatus, the magnetic field gradient is produced by 1 stack of five permanent magnets each of 6.4 x 6.4 x 2.5 mm^3 dimensions (26). The total magnetic field is approximately that of a solenoid with its long axis in the *z* direction thus the resulting force on a given bead in the sample is almost exclusively in the z direction, and

varies by less than 1% over the region of the liquid sample monitored in the experiment. The magnets were held in a lateral position with respect to the micro-channel on a 3-axis translation stage in order to exert a force perpendicular to the glass surface to which the DNA was bound. The magnitude of the force applied on the beads was determined by the distance between the magnet and the glass surface.

Control Experiments Control experiments are performed for every pairing experiment. In the controls, each of the two types of labeled DNA is incubated separately with the beads and inside the capillary. In such reactions, no molecule contains both Dig- and biotin labels and thus no bead can be specifically bound to both a bead and the surface; however, a low level non-specific binding does result in some tethered beads. In experiments that measured the comparing of two complete λ molecules, the background arises from the non-specific binding, with the non-specific binding of the biotin labeled DNA giving a higher background than the non-specific binding of the Dig-labeled DNA. In experiments that considered the pairing of shorter sequences with λ, the background arises primarily from nonspecific binding of the bead to the unlabeled end of the λ-Dig molecule, with the resulting signals found at a distance corresponding to fully-extended λ DNA (e.g. Fig. S2). Though the shorter DNA may also non-specifically bind at one end, their fully extended lengths are so short that they were not easily distinguished from beads that adhere directly to the surface of the capillary, which are not counted in the assay. In all of the presented experiments, the number of tethered beads in the pairing signal was more than five times the number in either control sample, with one exception as noted below.

Preparation of 5 kb fragments. Briefly, the fragments were amplified using Pfu Ultra II fusion (Stratagene, Carlsbad, Ca) in a thermocycler. Typical conditions were as follows: 1 ng λ DNA, 1 X buffer, 1 mM $MgCl_2$, 0.5 uM dNTPs, 1 Unit Pfu Ultra II fusion. The cycling protocol was 5 min at 95°C, 30 cycles of 30 sec at 95°C, 30 sec at 55°C, 3 minutes at 72°C, and 15 min at 72°C. The oligonucleotides had the following sequences: fragment 5 kb comprising bp 21613 to 26598 on λ DNA, 5'-XUUUGCTCATGCCCACACAAGTG-3' where X is biotin and the U's are 2'Ome RNA bases and reverse 5'-GAAAGCGTCCTTAACACCTC-3'. Fragment 5 kb comprising bp 1302 to 16322 on λ DNA, 5'-GGGCGGCGACCUCGTATGTTGCTCAGTTGCAG-3', where the first 12 bp are identical to a λ end and complementary to a biotinylated oligo (AGG TCG CCG CCC TTT Biotin 3') and the U is a 2'Ome RNA base, and reverse 5'-GCCATGTTGTTGCTGTATGC-3'. Fragment 5 kb comprising bp 116 to 5525 on λ DNA, 5'-XUUUAAACGACAGGTGCT GAAAGC-3' where X is biotin and the Us are 2'Ome RNA bases and reverse 5'-CCTCCTCACAGTTGAGGATC-3'. Pfu and many other proof reading enzymes do not copy 2'Ome RNA base pairs and thus an overhang is created during PCR where the sequence as well as the length of the sequence are predetermined by the designed oligonucleotide sequence. Following PCR, the fragments were separated via gel electrophoresis on a 1X TBE 1% agarose gel. The fragments were gel purified using a Nucleospin kit (Machery and Nagel, Bethlehem, PA).

1. Rau D, Parsegian VA (1992) Direct measurement of the intermolecular forces between counterion-condensed DNA double helices Evidence for long range attractive hydration forces. *Biophys J* 61:246-259

2. Cohen JA, Podgornik R, Hansen PL, Parsegian VA (2009) A Phenomenological One-Parameter Equation of State for Osmotic Pressures of PEG and Other Neutral Flexible Polymers in Good Solvents. *J Phys Chem B* 113:3709–3714.

3. Ramos JE, deVries R Jr., Neto JR (2005) DNA Ψ-condensation and reentrant decondensation: effect of the PEG degree of polymerization *J Phys Chem B*, 109: 23661–23665.

4. Kornyshev AA, Lee DJ, Leikin S, Wynveen A (2007) Structure and interactions of biological helices. *Rev Mod Phys* 79:943-996

5. Kornyshev AA, Leikin S (1999) Electrostatic zipper motif for DNA aggregation. *Phys Rev Lett* 82:4138-4141.

6. Kornyshev AA, Leikin S (2001) Sequence Recognition in the Pairing of DNA Duplexes. *Phys Rev Lett* 86:3666-3669.

7. Gelbart WM, Bruinsma RF, Pincus PA, Parsegian VA (2000) DNA-inspired electrostatics. *Physics Today* 53: 38-44

8. Grosberg AY, Nguyen TT, Shklovskii BI (2002) *Rev. Mod. Phys.* 74:329-345.

9. Dickerson RE (1992) DNA structure from A to Z. *Methods Enzymol* 211:67-111.

10. Olson WK, Zhurkin VB (2000) Modelling DNA deformations. *Curr Opin Struct Biol* 10:286-297.

11. Howerton SB, Sines CC, vanDerveer D, Williams LD (2001) Locating monovalent cations in the grooves of B-DNA. *Biochemistry*, 40:10023-10031

12. Chiu TK, Dickerson RE (2000) One angstrom crystal structures of B-DNA reveal sequence specific binding and groove-specific bending of DNA by magnesium and calcium. *J Mol Biol* 301: 915-945

13. Hud, N (2008) Nucleic Acid Metal Ion Interactions (The Royal Society of Chemistry, Cambridge, UK).

14. Leikin S, Parsegian VA (1994) Temperature-induced complementarity as a mechanism for biomolecular assembly. *Proteins* 19:73-76.

15. Cherstvy AG, Kornyshev AA, Leikin S (2004) Torsional Deformation of Double Helix in Interaction and aggregation of DNA. *J Phys Chem B* 108:6508-6518.

16. Bloomfield V A. (1996) DNA condensation. *Curr Opin Struct Biol* 6:334-341.

17. Hud, N V, Vilfan ID (2005) Toroidal DNA condensates: Unraveling the fine structure and the role of nucleation in determining size. *Annu Rev Biophys Biomol Struct* 34: 295-318.

18. Shklovskii BI (1999) Screening of a macroion by multivalent ions: Correlation-induced inversion of charge. *Phys Rev E* 60:5802-5811

19. Baldwin GS et al. (2008) DNA double helices recognize mutual sequence homology in a protein free environment *J Phys Chem* 112:1060-1064.

20. Zinchenko AA, Yoshikawa K (2005) Na^+ shows a markedly higher potential than K^+ in DNA compaction in a crowded environment. *Biophys J* 88:4118–4123.

21. Grønbech-Jensen N, Mashl RJ, Bruinsma RF, Gelbart WM (1997) Counterion-induced attraction between rigid polyelectrolytes. *Phys Rev Lett* 78:2477-2480.

22. Wynveen A, Lee DJ, Kornyshev AA, Leikin S (2008) Helical coherence of DNA in crystals and solution. *Nucleic Acids Res* 36:5540-5551.

23. McGavin S (1971) Models of specifically paired like (homologous) nucleic acid structures. *J Mol Biol* 55:293-298.

24. Johnson JE, Smith JS, Kozak ML, Johnson FB. (2008) In vivo veritas: using yeast to probe the biological functions of G-quadruplexes. *Biochimie* 90:1250-63.

25. Inoue S, Sugiyama S, Travers AA, Ohyama T (2007) Self-assembly of double-stranded DNA molecules at nanomolar concentrations *Biochemistry* 46: 164-171.

26. Danilowicz C et al. (2003) DNA unzipped under a constant force exhibits multiple metastable intermediates. *Proc Natl Acad Sci USA* 100:1694-1699.

Supporting Figure Legends

Fig. S1. Pairing of Dig-labeled λ DNA incubated with biotin-labeled fragments in which regions of homology are flanked by long non-homologous tails. (A) Gentle preparation of biotinylated DNA fragments comprising λ DNA plus flanking non-lambda sequences at each end (for details see SI Materials and Methods). In brief: the desired λ DNA

segment was inserted into pRS305. The series of λ segments used is shown in (B). For each segment, plasmid DNA was prepared in the absence of phenol extraction, digested with the indicated restriction enzymes and a biotinylated linker ligated onto one end. Between the digestion and ligation steps, and after ligation, DNAs were purified and concentrated through spin columns. DNA species were then separated on an agarose gel and extracted from the gel by electroelution (see Supplemental procedures). Note that the non-biotinylated end of the substrate fragment is blunt end (SfoI) to prevent Watson-Crick base pairing with another such molecule during the pairing reaction. Also, the bead is separated from its adjacent dsDNA terminus by a 3' ssDNA link that permits freer bead rotation. Lambda DNA segments ranged in size from 1-5 kb; flanking non-λ regions were the same 400 and 4,500 bp segments in all cases.

(B) DNA substrates containing λ DNA segments of various lengths from different positions of the λ sequence. Schematic diagrams are shown for constructs of 5 kb, 2 kb, 1 kb λ DNA containing flanking regions from pRS305 plasmid as described in (A).

(C) Distribution of extensions for pairing between Dig-labeled λ DNA incubated with biotinylated fragments without flanking tails (first row) and with long flanking non-homologous tails attached to homolog subregions of decreasing length: 5 kb (second row), 2 kb (third row), and 1 kb (fourth row). Positions of each subregions are detailed on each plot and correspond to # 21613, 11302, and 116, blue, green, and red/orange, respectively. The lavender outlines show the histogram for a sequence where the non-homologous tails are directly joined without any homolog subregion. The number of bound beads in the controls was more than an order of magnitude lower than the number bound for the homologs, but the control distribution does not peak at the length characteristic of λ. Experiments were performed under standard pairing reaction conditions analogous to those in Figure 4 (text). Distances were measured for about 50-100 tethered beads in each experiment.

Fig. S2. Pairing of 6µg/ml biotinylated 5 kb fragments and 60 µg/ml Dig-labeled λ phage molecules. (A) Biotinylated fragments were copied from λ phage subregion at position 21613 (blue) and pcDNA3.1 vector as template (light blue). (B). Histograms for the measurements after 1 h incubation period from the curves shown in (A). (C) The

pcDNA3.1 histogram shown in light blue in (B) is shown on a larger scale so that the position distribution is clear. The distribution is dominated by the length characteristic of a λ molecule and they probably correspond to λ molecules non-specifically bound to the beads.

Fig. S3. Pairing of Dig-labeled DNA samples 60 µg/ml in PBS buffer with biotinylated DNA molecules of different length: 60 µg/ml full λ DNA (red), 6 µg/ml 5 kb subregion of λ DNA position 21613 (blue), 1.2 µg/ml 1 kb subregion of λ DNA position 21613 (light blue), 6 µg/ml ~5 kb copied from pcDNA3.1 vector (brown).

Fig. S4. Pairing of 6 µg/ml biotinylated 5 kb fragments and 60 µg/ml Dig-labeled λ phage molecules in the absence (blue) and presence of non specific competitors: 50 µg/ml human genomic dsDNA (red), and 0.1 % m/v bovine serum albumin (BSA) (green).

Figure S1A.

Preparation of DNAs containing 1-5kb sequences from bacteriophage λ embedded in non-lambda sequences (further details in legend).

pRS305 + λ DNA

340bp   4.5kbp

λ DNA

NgoMIV / NotI   SalI   SfoI
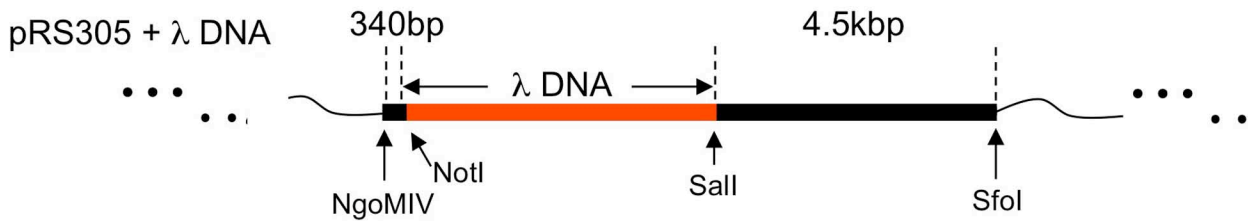
1. Preparation of plasmid DNA from cells

2. Creation of desired molecules

   (a) Restriction digestion of plasmid (NgoMIV + SfoI)

   (b) Ligation of Biotinylated linker onto NgoMIV site

   (c) DNA purification (G-50 spin column)

   (d) DNA concentration (YM-100 filter)

Biotinylated linker with 3'ssDNA link

5'Biotin 'AAA'

3'GGCC5'

340bp
400bp

4.5kbp

NotI
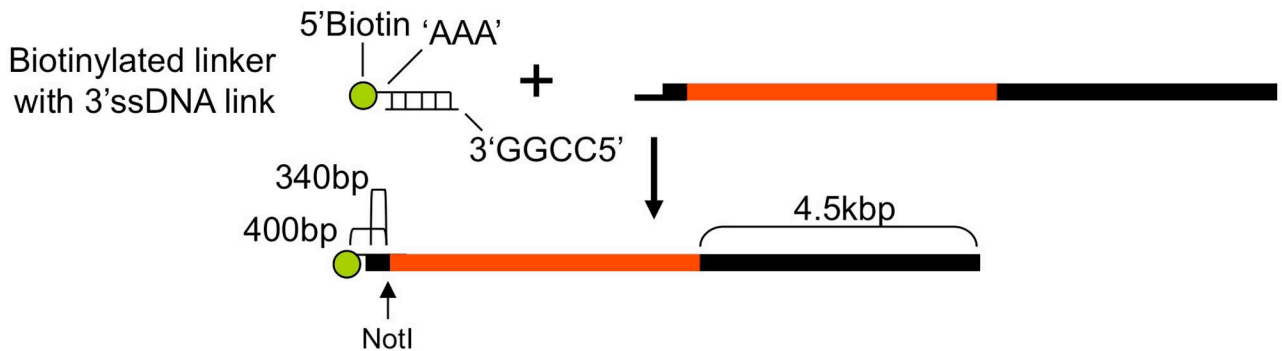
3. Preparation of molecule population

   (a) Gel electrophoresis (TBE; no ethidium bromide)

   (b) Electroelute DNA from gel (position marked by parallel EtBr-stained lane)

   (c) DNA purification (G-50 spin column)

   (d) DNA concentration (YM-100 filter)

   (e) Measure DNA concentration of aliquot (Nanodrop spectrophotometer).

4. Assess ligation efficiency in aliquot: NotI cleavage + 5' ^{32}P labeling

   plus gel electrophoresis distinguishes 340bp vs 400bp fragments corresponding,

   respectively, to absence vs presence of biotinylated linker

Figure S1B.

Long-tailed molecules analyzed for pairing with λ DNA in Fig. S4.

B. Molecules analyzed.

| Plasmid pKK	λ DNA sequence bp positions	(kb)
pKK146	21613-26598	(5kb)
pKK156	21613-23613	(2kb)
pKK155	21613-22613	(1kb)
pKK157	116-5525	(5.4kb)
pKK159	116-2116	(2kb)
pKK158	116-1116	(1kb)
pKK160	11302-16322	(5kb)
pKK162	11302-13302	(2kb)
pKK161	11302-12302	(1kb)
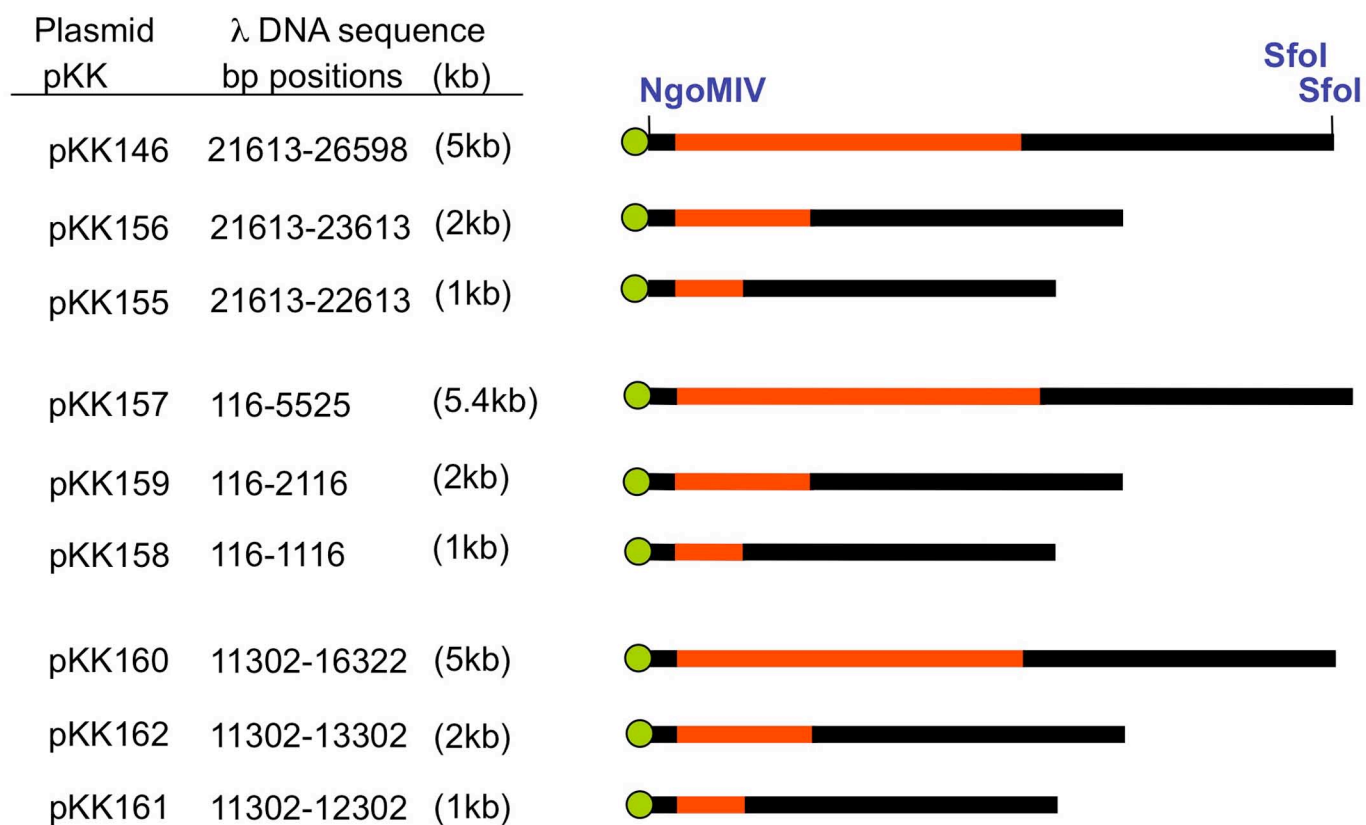
Figure S1C. Distributions of positions of tethered beads for different length homologous sequences with long flanking non-homologous tails
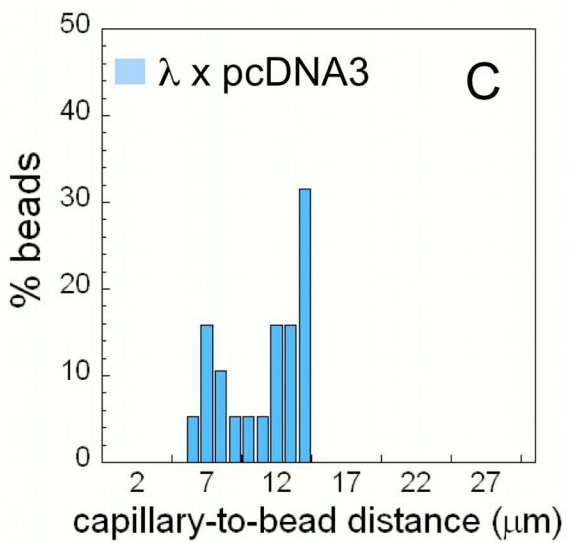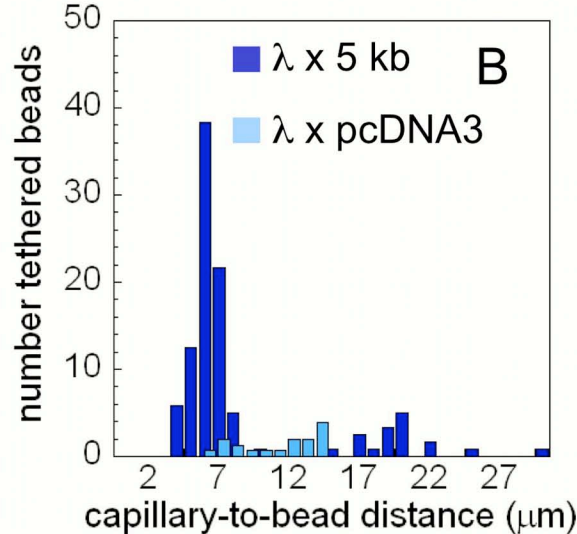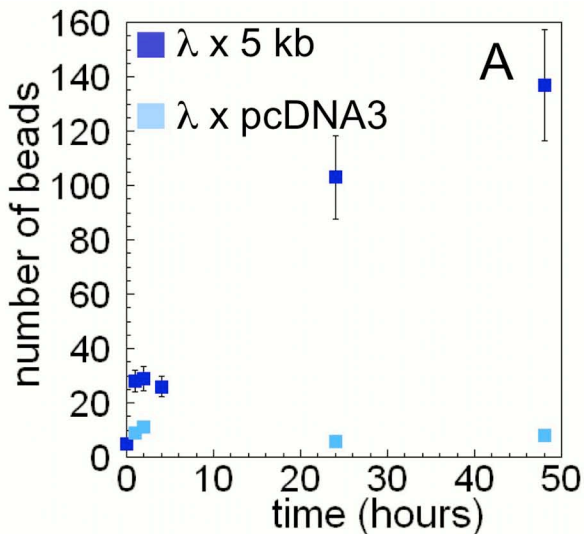
Untailed DNAs (from Fig. 3)
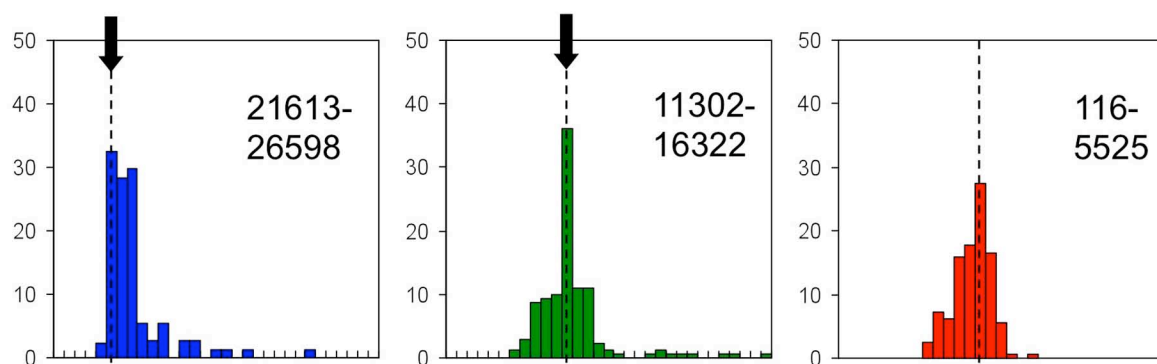
Long-tailed DNAs (described in Fig. S1B)

Percentage of tethered beads
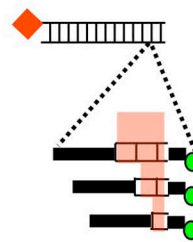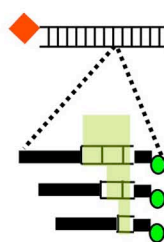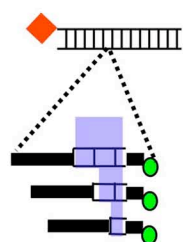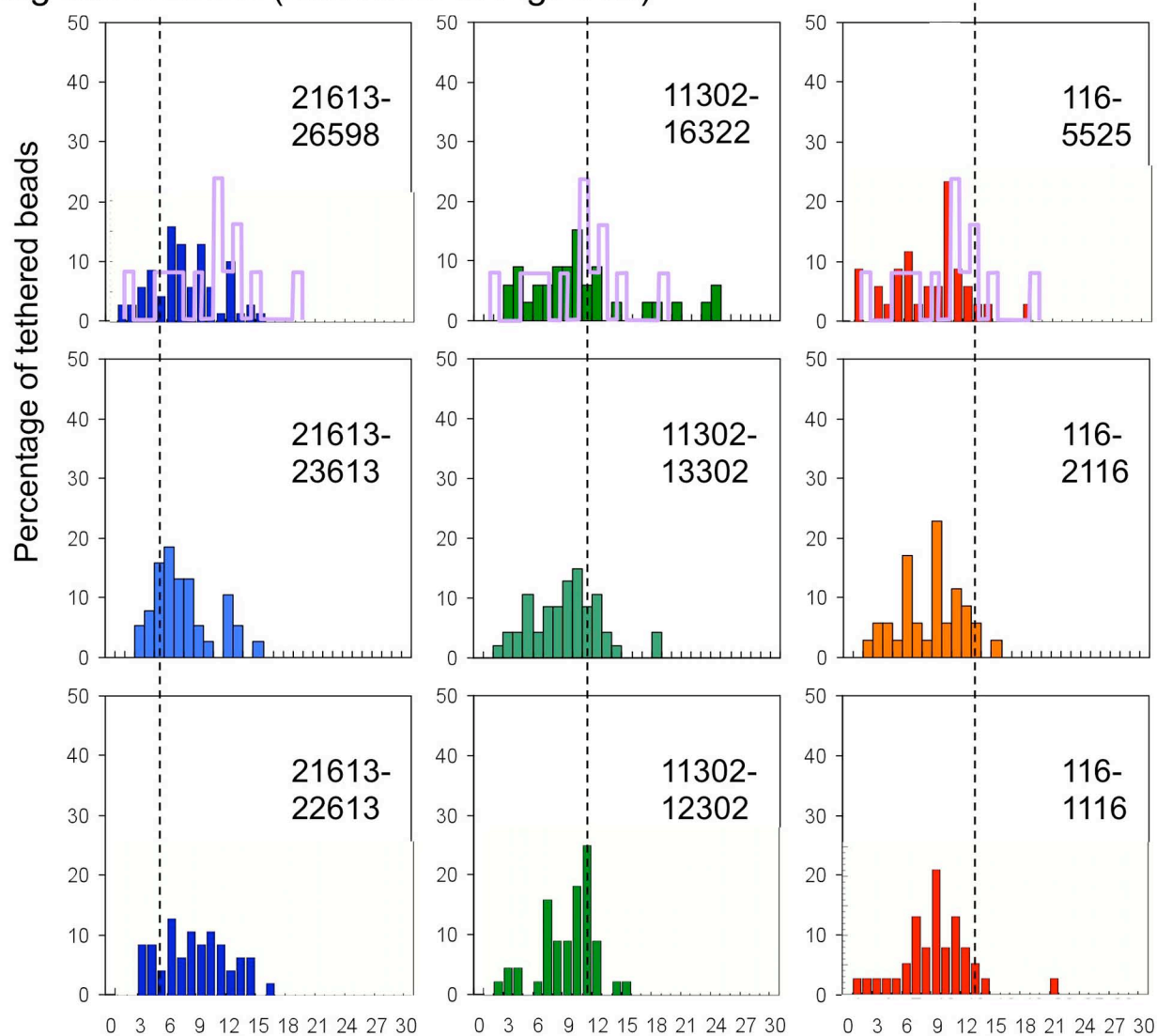
extension (um)
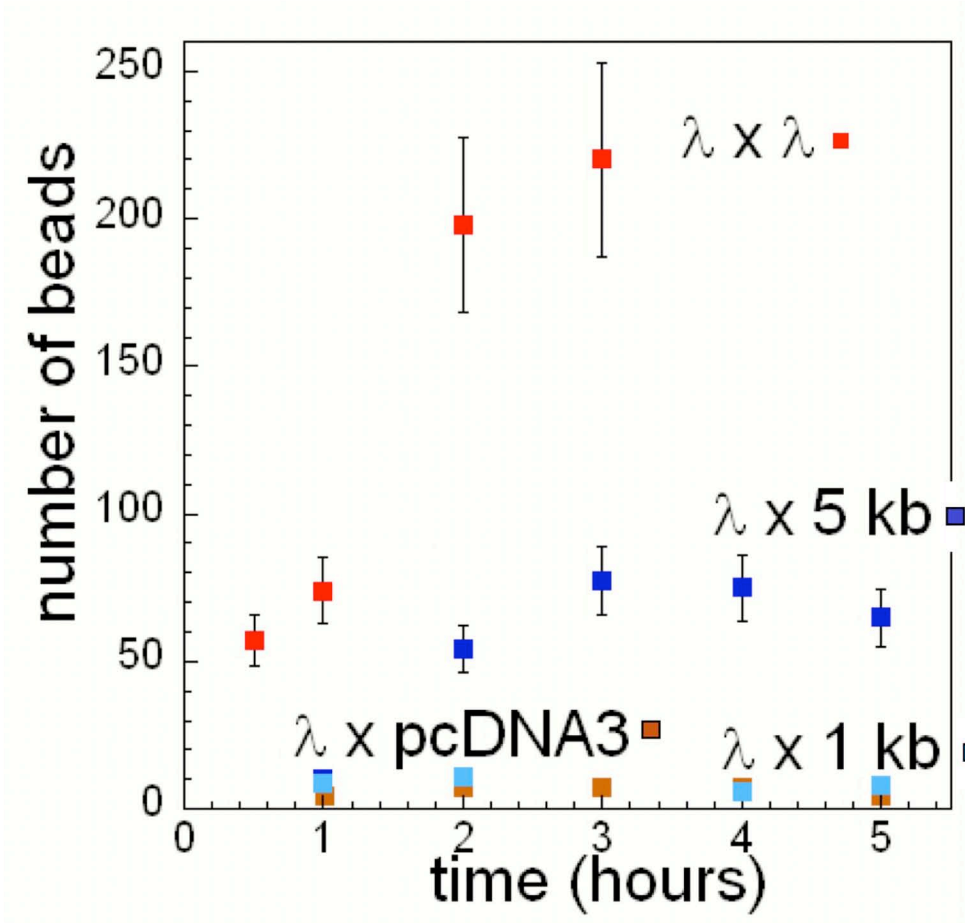
Figure S2.

λ DNA does not give a significant pairing signal with non-homologous DNA (pcDNA3.1)

Figure S3

Pairing of λ DNA with tail-less partners of decreasing length

Figure S4