

# DNA knots reveal a chiral organization of DNA in phage capsids

Javier Arsuaga<sup>\*†‡</sup>, Mariel Vazquez<sup>\*</sup>, Paul McGuirk<sup>§</sup>, Sonia Trigueros<sup>¶</sup>, De Witt Summers<sup>||</sup>, and Joaquim Roca<sup>¶\*\*</sup>

Departments of <sup>\*</sup>Mathematics, <sup>†</sup>Molecular and Cell Biology, and <sup>§</sup>Physics, University of California, Berkeley, CA 94720; <sup>||</sup>Department of Mathematics, Florida State University, Tallahassee, FL 32306; and <sup>¶</sup>Department of Molecular Biology, Institut de Biologia Molecular de Barcelona-Consejo Superior de Investigaciones Cientificas, 08034 Barcelona, Spain

Edited by Charles R. Cantor, Sequenom, Inc., San Diego, CA, and approved May 10, 2005 (received for review December 14, 2004)

Icosahedral bacteriophages pack their double-stranded DNA genomes to near-crystalline density and achieve one of the highest levels of DNA condensation found in nature. Despite numerous studies, some essential properties of the packaging geometry of the DNA inside the phage capsid are still unknown. We present a different approach to the problems of randomness and chirality of the packed DNA. We recently showed that most DNA molecules extracted from bacteriophage P4 are highly knotted because of the cyclization of the linear DNA molecule confined in the phage capsid. Here, we show that these knots provide information about the global arrangement of the DNA inside the capsid. First, we analyze the distribution of the viral DNA knots by high-resolution gel electrophoresis. Next, we perform Monte Carlo computer simulations of random knotting for freely jointed polygons confined to spherical volumes. Comparison of the knot distributions obtained by both techniques produces a topological proof of nonrandom packaging of the viral DNA. Moreover, our simulations show that the scarcity of the achiral knot  $4_1$  and the predominance of the torus knot  $5_1$  over the twist knot  $5_2$  observed in the viral distribution of DNA knots cannot be obtained by confinement alone but must include writhe bias in the conformation sampling. These results indicate that the packaging geometry of the DNA inside the viral capsid is writhe-directed.

bacteriophage | DNA condensation | DNA electrophoresis | Monte Carlo simulation | DNA writhe

All icosahedral bacteriophages with double-stranded DNA genomes are believed to pack their chromosomes in a similar manner (1). During phage morphogenesis, a procapsid is first assembled, and a linear DNA molecule is actively introduced inside it by the connector complex (2, 3). At the end of this process, the DNA and its associated water molecules fill the entire capsid volume, where DNA reaches concentrations of 800 mg/ml (4). Some animal viruses (5) and lipo-DNA complexes used in gene therapy (6) are postulated to hold similar DNA arrangements as those found in bacteriophages.

Although numerous studies have investigated the DNA packing geometry inside phage capsids, some of its properties remain unknown. Biochemical and structural analyses have revealed that DNA is kept in its B form (7–9) and that there are no specific DNA-protein interactions (10, 11) or correlation between DNA sequences and their spatial location inside the capsid, with the exception of the *cos* ends in some viruses (12). Many studies have found that regions of the packed DNA form domains of parallel fibers, which in some cases have different orientations, suggesting a certain degree of randomness (8, 9, 13, 14). The above observations have led to the proposal of several long-range organization models for DNA inside phage capsids: the ball of string model (13), the coaxial spooling model (8, 11, 13, 14), the spiral-fold model (15), and the folded toroidal model (16). Liquid crystalline models, which take into account properties of DNA at high concentrations and imply less global organization, have also been proposed (9). In this study, we present a different

approach to investigating the packaging geometry of DNA inside phage capsids.

The bacteriophage P4 has a linear, double-stranded DNA genome that is 10–11.5 kb in length and flanked by 16-bp cohesive *cos* ends (17). It has long been known that extraction of DNA from P4 phage heads results in a large proportion of highly knotted, nicked DNA circles (18, 19). DNA knotting probability is enhanced in P4 derivatives containing genome deletions (20) and in tailless mutants (21). We recently showed that most of these knots are formed by the cyclization of the DNA while it is confined inside the phage capsid (22). Here, we show that these knots capture information about the global arrangement of the packed DNA. By using high-resolution gel electrophoresis, we examine the knot distribution of the viral DNA. Next, we perform Monte Carlo computer simulations of random knotting for freely jointed polygons confined to spherical volumes and compare them with the experimental knot distributions. Our results demonstrate that long-range organization of DNA in the viral capsid is not random and indicate that the packing geometry of the DNA is writhe-directed.

## Materials and Methods

**Isolation of Knotted DNA from P4 Phage Particles.** Bacteriophage P4 *vir1 del22* was prepared as described in ref. 22. Briefly, the phage was amplified in *Escherichia coli* strain C-1895, which is lysogenic for P2 (the helper prophage). Stocks of P4 *vir1 del22* were used to infect *E. coli* C-8001, which is lysogenic for P2  $\Delta$ H13. Because gene H encodes part of the P4 phage tail, this strain produced the tailless mutant of P4 *vir1 del22*. DNA was obtained from phage particles by following the procedure described by Isaksen *et al.* (21) with minor modifications: After bacterial lysis, phages were precipitated with polyethylene glycol 8000. Phage precipitates were dissolved in phage buffer containing 10 mM  $MgCl_2$ /10 mM Tris-HCl, pH 7.2/130 mM ammonium acetate. Phage particles were banded by cesium chloride centrifugation in an NVT65 rotor (Beckman) for 14 h at 45,000 rpm. The banded particles were dialyzed against phage buffer and extracted twice with phenol and once with phenol/chloroform to obtain DNA. Phage DNA was dialyzed against TEN buffer (10 mM Tris-HCl, pH 8/1 mM EDTA/100 mM NaCl) and kept at 4°C.

**Electrophoretic Analysis of Knotted DNA.** Purified DNA was analyzed by two-dimensional gel electrophoresis as described by Trigueros *et al.* (23). Gel slabs of 0.4% agarose concentration that were equilibrated with TBE buffer (100 mM Tris-borate, pH 8.3/2 mM EDTA) were used. In the first dimension, DNA samples were run at 0.8 V/cm for 40 h at room temperature. After a 90° rotation of the gel, a second dimension was run in the same electrophoresis buffer at 3.4 V/cm for 4 h at room

This paper was submitted directly (Track II) to the PNAS office.

<sup>†</sup>Present address: Department of Mathematics and Center for Computation in the Life Sciences, San Francisco State University, San Francisco, CA 94132.

<sup>\*\*</sup>To whom correspondence should be addressed. E-mail: jrbbmc@cid.csic.es.

© 2005 by The National Academy of Sciences of the USA

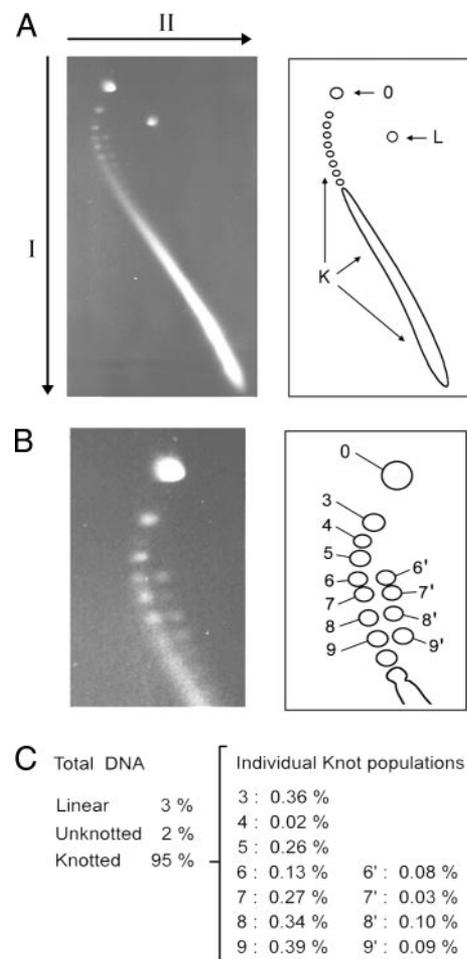
temperature. After ethidium bromide staining of DNA and photography, gels were blotted to nylon membranes, and the DNA was radioprobed for phosphorimaging quantification. A marker ladder for twist-type knots was generated by incubating a supercoiled 10-kb plasmid with equimolar amounts of T4 topoisomerase II, as described by Wasserman and Cozzarelli (24). After the knotting reaction, supercoils were removed from DNA by limited nicking in a reaction containing 50 mM Tris (pH 7.5), 5 mM MgCl<sub>2</sub>, 100 µg/ml each DNA and ethidium bromide, and 2 µg/ml DNase I.

**Knot Type Probabilities for P4 DNA in Free Solution.** The probability that a knot  $K$  of  $n$  statistical lengths and diameter  $d$  is formed in free solution is given by  $P_K(n, d) = P_K(n, 0)\exp(-rd/n)$ , where  $r$  depends on the knot type and equals 22 for the knot 3<sub>1</sub> and 31 for the knot 4<sub>1</sub> (25). The knotting probability of a 10-kb DNA molecule cyclized in free solution is  $\approx 0.03$  (25, 26), which implies an effective DNA diameter near 35 Å. Because  $P_{3_1}(34, 0) = 0.06$  and  $P_{4_1}(34, 0) = 0.009$ , then  $P_{3_1}(34, 35) = 0.027$  (1/36 times that of the unknot) and  $P_{4_1}(34, 35) = 0.003$  (1/323 times of the unknot). These values were used to estimate the fractions of the knot 3<sub>1</sub> and the knot 4<sub>1</sub> generated for P4 DNA in free solution.

**Monte Carlo Simulations.** Algorithms that generate random distributions of equilateral polygons with zero volume and confined to spherical volumes, as well as algorithms that compute the Alexander polynomial ( $\Delta(t)$ ), have been reported previously (22). Each selected knotted polygon was further identified by evaluating its Alexander polynomial at  $t = -2$  and  $t = -3$ . Although the Alexander polynomial is an excellent discriminator among knots of low crossing number and its computation is fast, it does not distinguish completely among some knotted chains [for example, composite knots 3<sub>1</sub>#3<sub>1</sub> and 3<sub>1</sub>#4<sub>1</sub> have polynomials identical to those of prime knots 8<sub>20</sub> and 8<sub>21</sub>, respectively (27)]. Evaluation of the polynomial at  $t = -2$  and  $t = -3$  is also ambiguous because the Alexander polynomial is defined up to units in  $Z[t^{-1}, t]$  (28), and therefore the algorithm returns  $(\pm n^{\pm m} \Delta(-n))$ , where  $n = 2$  or 3 and  $m$  is an integer. To deal with this uncertainty, we followed van Rensburg and Whittington (27) and chose the largest exponent  $k$  such that the product  $(\pm n^{\pm m} \Delta(-n))n^{\pm k}$  is an odd integer with  $n = 2$  or 3. This value was taken as the invariant. To compute the writhe, we generated >300 regular projections and diagrams for each selected polygon. To each of the projected crossings a +1 or -1 was assigned by the right-hand convention. The directional writhe for each diagram was computed by summing these values. The writhe was then determined by averaging the directional writhe over all possible projections (29). Computer programs were tested by comparing the results with those previously reported (30). To generate writhe-directed random distributions of polygons, we used a rejection method in which polygons whose writhe was below a positive value were not sampled.

## Results and Discussion

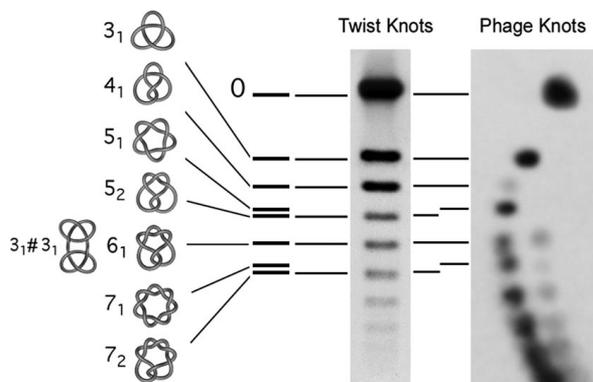
**Knot Complexity of DNA Molecules Extracted from Phage P4.** We extracted the 10-kb DNA from the tailless mutant of phage P4 *vir1 del22*, which produces 95% knotted molecules (22), and analyzed it by a high-resolution two-dimensional gel electrophoresis (23) (Fig. 1A). This technique allowed us to separate DNA knot populations according to their crossing number (i.e., the minimal number of crossings over all projections of a knot), as well as to separate some knot populations of the same crossing number (31, 32). In the first dimension (at low voltage), individual gel bands corresponding to knot populations having crossing numbers between three and nine were discernible; knots with higher crossing numbers were embedded in a long tail (denoted as K in Fig. 1A). The second dimension (at high voltage) further resolved individual gel bands corresponding to



**Fig. 1.** Analysis of knotted DNA by gel electrophoresis. (A) DNA was extracted from tailless mutants of phage P4 *vir1 del22* and analyzed by two-dimensional agarose gel electrophoresis as described in *Materials and Methods*. The first dimension at low voltage (top to bottom) separated DNA knot populations according to their crossing number. The unknotted DNA circle or trivial knot (0) has the slowest gel velocity, whereas knotted DNA populations (K) have gel velocity proportional to their crossing number. The second dimension at high voltage (left to right) segregated the linear DNA molecules (L) from the arched distribution of knotted molecules and further resolved some gel bands corresponding to knot subpopulations. (B) Upper area of the gel picture showing knot populations of low crossing number. Individual gel bands corresponding to knot populations containing three to nine crossings are indicated (labeled 3–9) in the main arch of the gel. A second arch of higher gel speed containing knot subpopulations of six and more crossings is generated by the second dimension of the electrophoresis. Individual gel bands of knot subpopulations of six to nine crossings (labeled 6'–9') are indicated. (C) Quantification of the individual knot populations of six to nine crossings (3–9 and 6'–9'). Both densitometric and phosphorimaging reading of three independent samples of DNA extracted from tailless mutants of phage P4 *vir1 del22* produced nearly identical results. The indicated percentage values are relative to the total amount of knotted molecules.

knot populations with crossing numbers between six and nine. Although knot populations containing three, four, and five crossings (denoted as 3–5 in Fig. 1B) migrated as single bands in a main arch of low gel velocity, knot populations containing six and more crossings split into two subpopulations (denoted as 6–9 and 6'–9' in Fig. 1B), creating a second arch of greater gel velocity.

We quantified the individual knot populations of three to nine crossings, which represented 2.2% of the total amount of knotted molecules (Fig. 1C). Densitometer readings confirmed the ap-



**Fig. 2.** Identification of specific knot types by their position in the gel. The gel velocity at low voltage of individual knot populations resolved by two-dimensional electrophoresis (Right) is compared with the gel velocity at low voltage of twist knots ( $3_1$ ,  $4_1$ ,  $5_2$ ,  $6_1$ , and  $7_2$ ) of a 10-kb nicked plasmid (Center) and with known relative migration distances of some knot types (refs. 30 and 31) (Left). Geometrical representations of the prime knots  $3_1$ ,  $4_1$ ,  $5_1$ ,  $5_2$ ,  $6_1$ ,  $7_1$ , and  $7_2$  and of the composite knot  $3_1\#3_1$  are shown. The unknotted DNA circle or trivial knot (0) is also indicated. Note that in the main arch of the two-dimensional gel and below the knots  $3_1$  and  $4_1$ , the knot population of five crossings matches the migration of the torus knot  $5_1$ , which migrates closer to the knot  $4_1$  than to the knots of six crossings. The other possible five-crossing knot, the twist knot  $5_2$ , appears to be negligible or absent in the viral distribution. Note also that the knot population of seven crossings matches the migration of the torus knot  $7_1$  rather than the twist knot  $7_2$ . In the secondary arch of the two-dimensional gel, the first knot population of six crossings has similar low-voltage migration of the composite knot  $3_1\#3_1$ .

parent scarcity of the knot of four crossings (knot  $4_1$ ) relative to the other knot populations in the main arch of the gel (denoted by 4 in Fig. 1B). It also made evident the shortage of the knot subpopulation of seven crossings in the second arch of the gel (denoted by  $7'$  in Fig. 1B). The scarcity of the knot  $4_1$  relative to the knot of three crossings (knot  $3_1$ ) and to other knot populations is enhanced if we make the correction for DNA molecules plausibly knotted outside the viral capsid. Namely, if a fraction of the observed knots is formed by random cyclization of DNA outside the capsid, then, in the worst case scenario, all observed unknotted molecules (no more than 5% of the total molecules extracted) will be formed in free solution. In such a case, one can predict that 38% of the total number of observed  $3_1$  knots and 75% of the observed  $4_1$  knots are formed by random knotting in free solution (25, 26). If all of the knots plausibly formed outside the capsid are removed from the observed knot distribution, the experimental values for knots  $4_1$  and  $3_1$  (1:18 ratio) would be corrected, resulting in a 1:44 ratio.

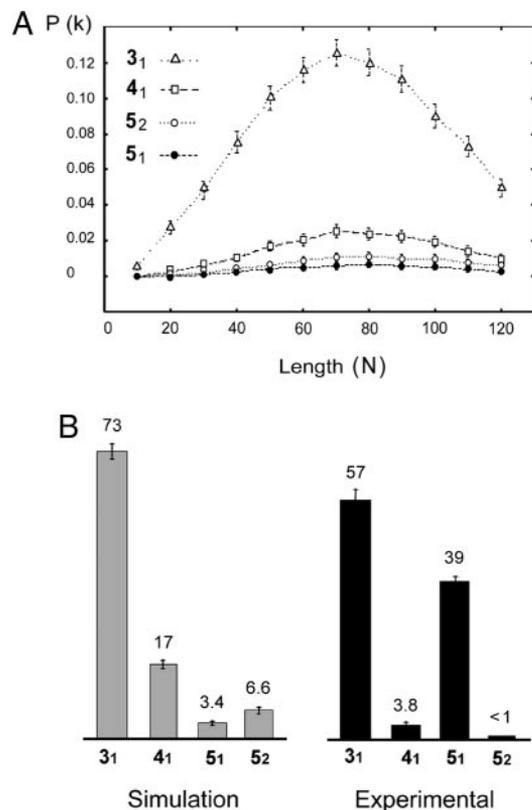
**Identification of Specific Knot Types by Their Location on the Gel.** Gel electrophoresis can distinguish some knot types with the same crossing number. For example, at low voltage, torus knots (such as  $5_1$  and  $7_1$ ) migrate slightly slower than their corresponding twist knots ( $5_2$  and  $7_2$ ) (31, 32). We used this knowledge in conjunction with a marker ladder for twist knots ( $3_1$ ,  $4_1$ ,  $5_2$ ,  $6_1$ , and  $7_2$ ) to identify several gel bands of the phage DNA matching the migration of known knot types (Fig. 2). In the main arch of the gel, in addition to the unambiguous knots  $3_1$  and  $4_1$ , the knot population of five crossings matched the migration of the torus knot  $5_1$ . The other possible five-crossing knot, the twist knot  $5_2$  that migrates between and equidistant to the four- and six-crossing knot populations, appeared to be negligible or absent. The knot population of seven crossings matched the migration of the torus knot  $7_1$  rather than the twist knot  $7_2$ , which has slightly higher gel velocity. Yet, we cannot identify this gel band

as the knot  $7_1$ , because other possible knot types of seven crossings cannot be excluded.

Several indicators led us to believe that the second arch of the gel consists of mainly composite knots (resulting from combinations of prime knots). First, the arch starts at knot populations containing six crossings, and no composite knots of fewer than six crossings exist. Second, the population of six crossings matched the migration at low voltage of the granny knot (composite of a  $3_1$  plus a  $3_1$ , indicated as  $3_1\#3_1$ ) (33), although the square knot (the other possible composite of six crossings,  $3_1\#-3_1$ ) cannot be excluded. Third, consistent with the low amount of  $4_1$  knots, the size of the seven-crossing subpopulation is also reduced; thus, any composite seven-crossing knot is  $3_1\#4_1$  (or  $-3_1\#4_1$ ). The increased gel velocity at high voltage (second gel dimension) of composite knots relative to prime knots of the same crossing number likely reflects distinct flexibility properties of the composites during electrophoresis.

**Monte Carlo Simulations of Random Knot Distributions in Confined Volumes.** Next, we asked whether the observed distribution of DNA knots could be compatible with a random embedding of the DNA inside the phage capsid. We used Monte Carlo simulations to model knotting of randomly embedded, freely jointed polygons confined to spherical volumes. Because the persistence length of the duplex DNA is not applicable in confined volumes (it is applicable in unbounded three-dimensional space), we considered freely jointed polygons as the zeroth approximation of the packed DNA molecule. Then, the flexibility of the chain is given by the ratio  $R/N$ , where  $N$  is the number of edges in the polygon and  $R$  is the sphere radius in edge-length units. When we computed random knot distributions for a range of chain lengths confined to spheres with a fixed radius, the probabilities of the knots  $3_1$ ,  $4_1$ ,  $5_1$ , and  $5_2$  produced nonintersecting distributions, with simpler knots being more probable (Fig. 3A). That is, the knot  $3_1$  is more probable than the knot  $4_1$ , and both are more probable than any five-crossing knot. In addition, the probability of the twist knot  $5_2$  is higher than that of the torus knot  $5_1$ . Similar results had been observed for other random polymer models with/without volume exclusion and with/without confinement (34–38), indicating that this phenomenon is model-independent. All of the simulated distributions, showing the monotonically decreasing amounts of knotted products with increasing crossing number, highly contrasted with our experimental distribution, in which the probability of the knot  $4_1$  is markedly reduced and in which the knot probability of the knot  $5_1$  prevails over that of the knot  $5_2$  (Fig. 3B). These differences provide a compelling proof that the embedding of the DNA molecule inside the phage capsid is not random.

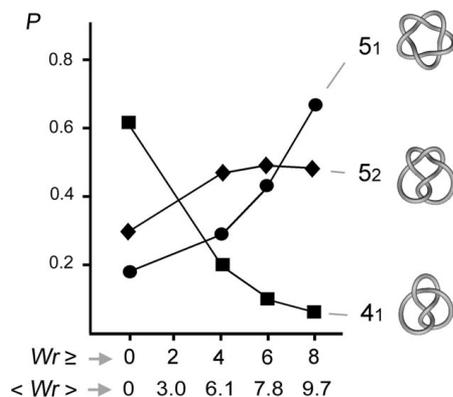
**Monte Carlo Simulations of Writhe-Directed Knot Distributions in Confined Volumes.** How can we explain the scarcity of  $4_1$  in the spectrum of viral knots? The knot  $4_1$  is achiral (equivalent to its mirror image). Random polygonal realizations of the  $4_1$  knot in free space and in confined volumes produce a family of polygons whose writhe distribution for any polygonal length is a Gaussian curve with zero mean (the writhe is a geometrical quantity measuring the signed spatial deviation from planarity of a closed curve) and whose variance grows as the square root of the length (29). Therefore, we argue that the main reason for the scarcity of the knot  $4_1$  is a writhe bias imposed on the DNA inside the phage capsid. To test this hypothesis, we simulated polygons randomly embedded in spheres whose mean writhe value was gradually increased. To induce writhe in the sampling, we used a rejection method in which polygons of writhe below a cutoff value were not sampled. Then, we calculated the probabilities of the prime knots  $4_1$ ,  $5_1$ , and  $5_2$  for each writhe-biased sampling. The results shown in Fig. 4 were computed with a freely jointed



**Fig. 3.** Comparison of experimental and computer-simulated distributions of knots. (A) Distribution probabilities ( $P(k)$ ) obtained by Monte Carlo simulations of the prime knots  $3_1$ ,  $4_1$ ,  $5_1$ , and  $5_2$  for closed ideal polymers of variable chain lengths ( $n$  = number of edges) confined to a spherical volume of fixed radius ( $R = 4$  edge lengths). Error bars represent standard deviations. (B) Comparison of the computed probabilities of the knots  $3_1$ ,  $4_1$ ,  $5_1$ , and  $5_2$  (for polymers of length  $n = 90$  randomly embedded into a sphere of radius  $R = 4$ ) with the experimental distribution of knots. The relative amount of each knot type is plotted. Note that fractions of knots  $3_1$  and  $4_1$  plausibly formed in free solution are not subtracted from the experimental distribution. If these corrections are considered, the relative amount of knot  $4_1$  is further reduced.

chain of 90 edges confined in a sphere of radius of 4 edge-length units. A drop of the probability of the knot  $4_1$ , as well as an exponential increase of the probability of the knot  $5_1$  but not of the knot  $5_2$ , readily emerged by increasing the writhe rejection value. The same results, but with knots of opposite sign, were obtained for knot distributions with the corresponding negative writhe values. These writhe-induced changes in the knot probability distribution are independent of the number of edges in the equilateral polygon and the sphere radius length. Accordingly, previous studies had shown that the mean writhe value of random conformations of a given knot does not depend on the length of the chain but only on the knot type and that these values are model-independent (39–42). Because the writhe-directed simulated distributions approach the observed experimental spectrum of knots, we conclude that a high writhe of the DNA inside the phage is the most likely factor responsible for the observed experimental knot spectrum.

Consistent with the involvement of writhe in the DNA packing geometry, it is also the reduced amount of prime knots of six crossings visible in the main arch of the gels (Fig. 1C). All prime knots of six crossings have a lower mean writhe value  $Wr$  ( $Wr$  of  $6_1 = 1.23$ ,  $Wr$  of  $6_2 = 2.70$ , and  $Wr$  of  $6_3 = 0.16$ ) than the torus knots of five and seven crossings ( $Wr$  of  $5_1 = 6.26$  and  $Wr$  of  $7_1 = 9.15$ ). In contrast, the negligible amount of the twist knot of five crossings ( $Wr$  of  $5_2 = 4.54$ ) in the experimental



**Fig. 4.** Effect of a writhe-biased sampling on the probability of knots  $4_1$ ,  $5_1$ , and  $5_2$ . The writhe of polygons of length  $n = 90$  randomly embedded into a sphere of radius  $R = 4$  were computed, and only conformations whose writhe values were higher than a prefixed value ( $Wr \geq 4, 6$ , or  $8$ ) were sampled. The computed mean writhe value ( $\langle Wr \rangle$ ) of each sampled population is indicated. The ratios of the probabilities of the knots  $4_1$ ,  $5_1$ , and  $5_2$  relative to that of the knot  $3_1$  for each writhe-biased sampling are plotted ( $P$ ).

distributions is striking. The apparent predominance of torus knots ( $5_1$  and  $7_1$ ) over twist knots ( $5_2$  and  $7_2$ ) in the experimental distribution suggests that writhe emerges from a toroidal or spool-like conformation of the packed DNA. Consistent with our findings, theoretical calculations of long-range organization of DNA by Monte Carlo (43) and molecular dynamics methods (44–47) favor toroidal and spool-like arrangements for DNA packed inside the phage capsids. Calculations of optimal spool-like conformations of DNA in phage P4 already predicted a large nonzero writhe (45). These studies gave an estimated writhe of  $-45$  for the 10-kb DNA, which closely corresponds with the level of supercoiling density typically found in bacterial chromosomes (45).

The actual writhe value of the DNA packaged in the phage P4 capsid cannot be estimated in the present study. The phage P4 capsid has a diameter of 38 nm. If the parameters used to compute writhe-biased ensembles as in Fig. 4 ( $n = 90$  and  $R = 4$ ) were applied to a 10-kb DNA molecule, they would translate into 90 segments of 35 nm confined in a model capsid of radius 140 nm. Likewise, our study cannot argue for or against recent models that suggest that to minimize DNA bending energy, a spool conformation might be concentric rather than coaxial (47). Therefore, beyond the main conclusion of this work that the distribution of viral knots requires the mean writhe of the confined DNA be nonzero, the applicability of our simulations to other aspects of the DNA packaging in phage P4 is limited. We argue that further identification of the knotted DNA populations will provide more critical information for the packing geometry of DNA inside the phage.

## Conclusions

Knots can be seen as discrete measuring units of the organizational complexity of filaments and fibers. Here, we show that knot distributions of DNA molecules can provide information on the long-range organization of DNA in a biological structure. We chose the problem of DNA packing in an icosahedral phage capsid and addressed the questions of randomness and chirality by comparing experimental knot distributions with simulated knot distributions. The scarcity of the achiral knot  $4_1$  and the predominance of the torus knot  $5_1$  in the experimental distribution highly contrasted with simulated distributions of random knots in confined volumes, in which the knot  $4_1$  is more probable than any five-crossing knot, and the knot  $5_2$  is more probable the knot  $5_1$ . To our knowledge, these

