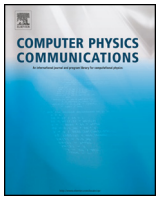
Computer Physics Communications 286 (2023) 108639

Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc



A computational package for measuring Topological Entanglement in Polymers, Proteins and Periodic systems (TEPPP) ☆,☆☆

Tom Herschberg [a], Kyle Pifer [b], Eleni Panagiotou [c,*,1]

[a] Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, 61801, IL, USA
[b] Department of Computer Science, University of Tennessee at Chattanooga, Chattanooga, 37403, TN, USA
[c] Department of Mathematics and SimCenter, University of Tennessee at Chattanooga, 613 McCallie Ave, Chattanooga, 37403, TN, USA

ARTICLE INFO

Article history:
Received 8 November 2021
Received in revised form 20 October 2022
Accepted 9 December 2022
Available online 13 January 2023

Keywords:
Software
Topology
Polymers
Proteins
Knots

ABSTRACT

Many materials, like polymer melts, solutions, biopolymers and textiles, are composed of entangled filaments. The entanglement in these systems significantly affects their mechanical properties and their function. We introduce the Topological Entanglement in Polymers, Proteins and Periodic systems (TEPPP) software, that enables to measure the topological and geometrical complexity in such systems. In particular, this software enables the computation of the Writhe, the Gauss linking integral and the Jones polynomial of each filament or pair of filaments in the system, whether they are open or closed. In particular, for systems employing Periodic Boundary Conditions (PBC), the software also allows to compute the total pairwise entanglement in PBC, using the Periodic linking number and the Periodic Writhe. For linear (open) chains, TEPPP can calculate all these topological parameters (including the Jones polynomial) without any closure scheme. In addition, TEPPP also enables measuring self and pairwise entanglement at different length-scales along a chain or a pair of chains. With appropriate preprocessing of input files, the code can also be used for star or branched architectures. We provide examples of how the code is used and we present results on the entanglement effect in polymers obtained using this package. We show how TEPPP can be used to measure the topological entanglement of linear polymer chains in a melt, revealing subtle entanglement transitions never seen before. We also used TEPPP to analyze the effect of knotting and its location in diblock copolymer melts, which reveals that knotting localization transition coincides with lamellar-disorder transition in these systems. Finally, we use TEPPP to reveal some of the topological structure of the SARS-CoV-2 Spike protein, which points to interesting structure in a region that contains the S1/S2 cleavage site.

Program summary

Program Title: Topological Entanglement in Polymers, Proteins and Periodic systems (TEPPP) software
CPC Library link to program files: https://doi.org/10.17632/ygdbpnhpzw.1
Developer's repository link: https://github.com/TEPPP-software
Licensing provisions: BSD 3-clause
Programming language: C++
Supplementary material:
Nature of problem: Measuring single and pairwise entanglement and knotting in systems of linear or ring filaments (open or closed curves) in 3-space or in systems employing Periodic Boundary Conditions (PBC) at different length scales.
Solution method: TEPPP can be used to measure topological entanglement complexity in single or multi-chain filament systems in 3-space or in systems employing PBC. Given as input the coordinates of the curves, TEPPP can compute the Gauss linking integral, the Writhe, the Jones polynomial, the Periodic

☆ The review of this paper was arranged by Prof. Blum Volker.

☆☆ This paper and its associated computer program are available via the Computer Physics Communications homepage on ScienceDirect (http://www.sciencedirect.com/science/journal/00104655).

* Corresponding author.
E-mail address: Eleni.Panagiotou@asu.edu (E. Panagiotou).
[1] Now at School of Mathematical and Statistical Sciences, Arizona State University, 900 Palm Walk, Tempe, 85281, AZ, USA.

https://doi.org/10.1016/j.cpc.2022.108639

Linking Number and the Periodic Writhe. Also, TEPPP can measure effects of local linking and knotting using scanning methods along the chains.

1. Introduction

Entanglement of collections of filaments arises in many physical systems, such as polymers, biopolymers, textile weaves and other woven structures. The entanglement of the constituent filaments in these systems affects their mechanical properties and their function [13,14,39,46]. Providing a model for the mechanical properties or function of these systems requires understanding of entanglement effects which in turn requires measuring the entanglement present. In this paper we introduce a new software that can be used to measure single or pairwise entanglement complexity in both open or closed, single or many-chain systems in 3-space and in Periodic Boundary Conditions (PBCs).

The current state-of-the art for measuring entanglement in many-chain systems, like polymer melts, is by using contour reduction algorithms [16,44,26]. These algorithms deform a polymer melt configuration in order to obtain a network from which local contacts can be derived. The number of these local contacts and the distance between them gives measures of entanglement complexity relevant to mechanics. However, as suggested by Edwards [13,14], entanglement is not a local contact, but rather a global constraint on a chain motion, as well as a multi-chain problem, which could be addressed using tools from topology [13,14]. In the 50 years since, however, only a limited number of studies of the topological entanglement of polymers have appeared. One of the main reasons is a gap in the mathematical progress of topology of open curves in 3-space, which has advanced only recently. Another important obstacle has been the difficulty of computing topological parameters for physical systems, both due to the technical difficulty of the algorithms, as well as due to the lack of communication of these concepts across disciplines. TEPPP fills this gap, by employing the most rigorous topological measures of entanglement of both linear or ring filaments that can enable a larger scientific community to use mathematics to advance knowledge in polymer physics, biopolymers and other filamentous systems.

Under some conditions, we can represent physical filaments by curves in 3-space. The entanglement of curves in 3-space is studied in Topology and, in particular, knot theory [1]. A knot is a simple closed curve in 3-space and a link is a collection of simple closed curves in 3-space. Essential to knot theory is the notion of topological equivalence, which can be briefly summarized by saying that two knots (resp links) are equivalent if we can go from one to the other without cutting and pasting. Topological invariants are functions defined on knots/links that are invariant under such deformations and can distinguish knots and links that are not equivalent (to various degrees of success, depending on the invariant). This notion is very useful if we are interested only in finding knots in our systems. However, the notion of entanglement in physical systems goes beyond knot/link type characterization. Entanglement in physical systems is closely related to time and length, two parameters irrelevant in traditional knot theory. For example, characteristics such as the tightness of a knot within a larger curve or the entanglement of open curves and the knottedness of open arcs are important in applications [13,14,28,42,11,20,18,19,45].

All of the existing packages aiming at measuring entanglement in filamentous structures are focused on measuring entanglement in proteins [41,40,9,22,25,27,10]. The only software to our knowledge which is applicable to many-chain systems, such as polymer melts or solutions, is given in [43]. However, all of this existing software deal with the open arcs by applying an artificial closure so that one obtains a knot for which topological invariants can be computed. Indeed, in the recent years, topological tools for biopolymers, in particular, have attracted a lot of attention and have been proven successful, especially in finding knots in proteins [41,40,9,22,25,27,10] or in finding knots in single chains in general [43]. The closure of the arcs however is problematic, not only due to the lack of mathematical rigor in this approximation, but also since it overly characterizes linear chains as trivial. For example, almost 90% of proteins are characterized as trivial using this method. Very importantly, when we look at molecular trajectories, this approach gives a discontinuous measure of complexity. Similar issues appear in approximating the curve in 3-space by a projection of it (a 2-dimensional object) [8].

In this manuscript, we introduce a software to analyze self- and pairwise entanglement in systems of single or many chains without employing any closure scheme. The software computes the Gauss linking integral [17] and the Jones polynomial of open and closed curves in 3-space [23,24,30]. The Gauss linking integral is a measure of self- or pairwise entanglement that is applicable to open and closed chains (linear or ring polymers) without any closure scheme [32,31,33,34]. For two curves, it is a continuous function of the chain coordinates and tends to an invariant of the link when the chains close. For one curve, the Gauss linking integral gives the Writhe, which is a continuous function of the chain coordinates and not invariant, even for closed curves. In [30], a new definition of the Jones polynomial was introduced, that showed that it can be applied to linear chains to give a polynomial with real coefficients which are continuous functions of the chain coordinates and it tends to the Jones polynomial invariant as the endpoints of the chain tend to coincide. All measures quantify the topological complexity of filaments in 3-space, but while the Writhe is more sensitive to the local geometry, the Jones polynomial is more sensitive to the knotting of the curves. These tools give a mathematically rigorous definition of entanglement without requiring an approximation of the open curves by knots. Moreover, these measures can be applied to open curves without open knots or slipknots and at multiple length scales to measure their complexity [34,32,31,33,2,21]. Since these are continuous functions of the chains, these measures are appropriate for studying the entanglement evolution of a system in time.

In some cases, especially for systems of single filaments, like proteins, the filament conformations are available from experiments [7,37]. However, in many cases, it is possible to see filament conformations only through molecular simulations. When using molecular simulations, in order to avoid boundary effects, PBCs are used [38]. PBCs arise also naturally in textiles or in the study of MOFS and zeolites [15,46,12]. Entanglement in PBC is more complex than in 3-space. None of the existing packages deals with measuring the total entanglement in periodic systems. The software introduced in this manuscript enables measuring the total pairwise entanglement in PBC through the application of the periodic linking number and the periodic writhe [35,29]. The periodic linking number and periodic writhe take into account the linking of multiple images of a chain in a periodic system and provide a continuous measure of entanglement that gives a topological invariant in the case of closed curves.

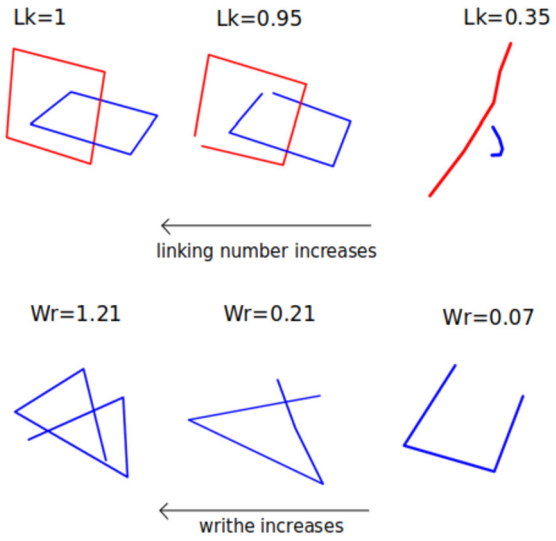


Fig. 1. The Gauss linking integral and the writhe can be applied to both open and closed curves in 3-space. For closed curves, the Gauss linking integral is a topological invariant that takes only integer values. For open curves, the Gauss linking integral is a real number that is a continuous function of the chain coordinates in 3-space. In both cases, the absolute value of the Gauss linking integral on average increases with increasing linking complexity. The writhe of both open and closed curves is not a topological invariant and it is a continuous function of the chain coordinates. The absolute value of the Writhe on average increases with increased self-entanglement of a curve. Figure from [30].

The paper is organized as follows: Section 2 describes the specifications of the TEPPP package and how it can be used. Section 3 shows examples of calculations done using the TEPPP package. Section 4 presents our conclusions. Appendix A describes in more detail the measures of topological entanglement used in the package and their properties.

2. The Topological Entanglement in Proteins, Polymers and Periodic structures (TEPPP) package

The Topological Entanglement in Proteins, Polymers, and Periodic structures (TEPPP) package is written in C++ with the ability to be run serially or in parallel using any MPI implementation. We stress that TEPPP can be used to measure the average pairwise or single-chain entanglement and knotting in a system of many chains. TEPPP takes as input ring or linear chains, but the measures can also be applied to other architectures, such as stars or branched chains, with appropriate pre-processing of input files. In this section we describe the features included in TEPPP and give a guide to its usage.

2.1. TEPPP features

TEPPP has the ability to compute the Gauss linking integral, the Writhe and the Jones polynomial for both closed and open curves in 3-space and in systems employing PBCs. The mathematical definitions of these measures are given in Appendix A.

2.1.1. Gauss linking integral and Writhe

The Gauss linking integral is a measure of linking of two chains and it gives an integer topological invariant in the case of closed chains, whose absolute value increases with link complexity. For a pair of open chains, it is a real number that is a continuous function of the curve coordinates and it also increases in absolute value with increasing pairwise entanglement. TEPPP can also compute the Writhe of a chain, which is the Gauss linking integral over one component. This gives a real number that is a continuous function of the chain coordinates for both open and closed chains and whose absolute value increases with increasing complexity of the conformation of the chain in 3-space. See Fig. 1 for an illustrative example.

The calculation of the Gauss linking integral and the Writhe of any pair of chains or any chain, respectively, is done using the algorithm described in [4] (see also Appendix A). With this algorithm, one obtains an exact closed form of the integral, avoiding numerical integration and using only dot, cross products and trigonometric functions. For a polymer chain with n monomers (or a polygonal chain with n vertices in general), the computation time scales as $O(n^2)$.

2.1.2. The Jones polynomial

Even though the Writhe can be helpful in assessing the complexity of a curve, it is not as useful in detecting knotting of a curve in 3-space, because it can be affected significantly by local geometrical features of the chains. The Jones polynomial is a stronger measure of knotting.

For closed curves (such as ring polymers), it gives a polynomial with integer coefficients, which is a topological invariant that classifies curves in knot types [23,24]. The Jones polynomial can in fact identify (at least in most practical cases) whether a ring polymer is a knot other than the unknot and whether two knots are the same. The knotting state of a system of closed curves cannot change without breakage of the chains. On average, the span of the polynomial (i.e. the range of its powers) or the highest power of the polynomial, increase with the complexity of the knot.

In [30] it was shown that the Jones polynomial can be well defined for open chains without any closure scheme. When applied to open chains (linear polymers) the Jones polynomial is a polynomial with real coefficients that are continuous functions of the chain coordinates [30]. See Fig. 2 for an illustrative example.

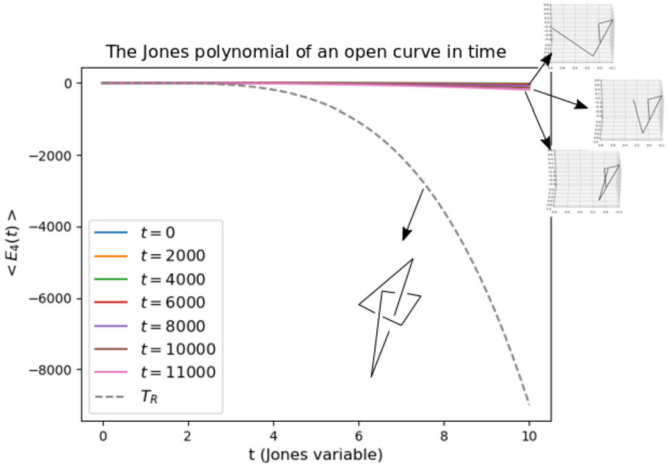


Fig. 2. The Jones polynomial of a polygonal curve in 3-space as it deforms in time to tighten a compact configuration, which is almost planar, resembling the knotoid k2.1. Initially, the Jones polynomial is $0.06t - 0.06t^{5/2} + 0.06t^{3/2} + 0.94$ and it tightens to a polynomial $0.71t - 0.71t^{5/2} + 0.71t^{3/2} + 0.29$. The dotted curve shows the Jones polynomial of the trefoil knot for comparison (which is $t + t^3 - t^4$). Higher powers and higher coefficients on average reflect higher knotting of a curve in 3-space. (For interpretation of the colors in the figures, the reader is referred to the web version of this article.)
Figure from [30].

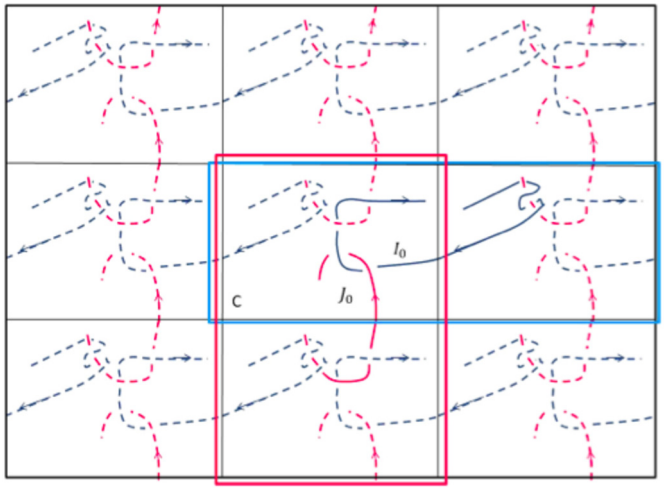


Fig. 3. Illustrative example of two chains in a system with 2 PBC. We define the periodic linking number as the sum of linking numbers between and image of chain I, I_0, (blue chain) and all the images of chain J (red chain), that intersect the cells in which the image I_0 of I unfolds. The periodic linking number is an integer topological invariant for closed chains in PBC and it is a continuous function of the chain coordinates in the case of open chains in PBC.

The Jones polynomial of a closed curve can be computed from any projection of a curve using the recursive formula shown in Eq. (A.3) in Appendix A. However, for open curves, the calculation requires averaging over all possible projection directions. To generate random projection directions, we generate random vectors using the uniform distribution on the unit sphere. In practice, 100 projections are sufficient to measure the topological complexity of an open curve in 3-space using the Jones polynomial.

The Jones polynomial is computed using a recursive calculation on the crossings of a diagram. The memory required for the calculation of the Jones polynomial of a chain diagram depends on the number of crossings, c, in the diagram and scales as $O(2^c)$. To reduce the computation time, we apply crossing reduction moves on a diagram, without altering the topology. This can be achieved through application of Reidemeister moves, RI, RII and RIII [1]. In TEPPP, one round of RI and RII is applied on each diagram, while respecting knotoid forbidden moves.

2.1.3. Entanglement in systems employing Periodic Boundary Conditions

Systems of filaments employing PBC generate an infinite system of curves in space. Computing the total linking in the system is impossible, as it is infinite and it contains infinitely many repetitions. On the other hand, the linking of the arcs in the simulation cell alone, cannot reflect the topological complexity these generate in 3-space. To address this, the Periodic Linking Number and the Periodic Writhe were introduced in [29] and it was shown that they provide well defined measures of linking, which do not account for any repetitions and capture the total linking in a periodic system. For systems employing PBCs, the TEPPP package allows to compute the Periodic Linking Number and the Periodic Writhe, which are sums of linking contributions between a parent image and the periodic images of another chain or the periodic images of itself, respectively (see Fig. 3 for an illustrative example). For closed chains, the Periodic Linking number and the Periodic Writhe are integers. For open chains, they are continuous functions of the chain coordinates whose absolute value increases with increasing pairwise linking of periodic images of other chains or of periodic images of the same chain,

Table 1
The commands included in the TEPPP package and their descriptions. Shown are the commands for serial execution. For execution in parallel, an "_mpi" extension is required for each command.

command	description
wr	Writhe of each chain in the system
wr_scan	Writhe of each subchain in the system
periodic_wr	Periodic Writhe of each chain in the system
lk	Gauss linking number (Lk) between each pair of chains in the system
lk_scan	Lk between each pair of subchains in the system
periodic_lk	Periodic linking number, LK_P, between each pair of chains in the system
jones	Jones Polynomial of each chain in the system
jones_scan	Jones Polynomial of each subchain in the system

respectively. These parameters can allow one to assess how much linking the periodic system imposes in total to the parent image of each chain.

2.1.4. Topology of subchains

The TEPPP package also provides the possibility to measure topological complexity of the chains at multiple length scales. Namely, the Writhe, the Gauss linking integral and the Jones polynomial can be computed at subsegments of the filaments of desired length using a sliding window method approach. This is done using the "scan" specifications. Using these specifications one can identify the tightness of knotted or linked structures and the location of entanglements.

2.2. TEPPP structure, requirements and usage

TEPPP is available for download at https://github.com/TEPPP-software. It includes a README file which contains all the required information about the package (including commands, input and output format requirements). The only dependency that TEPPP requires is a compiler that is compatible with C++17 or higher (GCC 10 and up). To build the serial/parallel or both versions of the software, run the following commend from within the "TEPPP" directory: `make serial` , `make parallel` , or `make all` , respectively.

Data files used with TEPPP commands must have the .teppp format and extension. The .teppp format is a list of the coordinates of the vertices of the constituent chains (thus, all input chains must be simple polygonal curves in 3-space, in unwrapped coordinates if in PBC). For example, for a system of 100 chains of length 20, a .teppp file contains 2000 coordinates, where the first 20 belong to chain 1, the next 20 to chain 2, etc. If the data file is in a format other than .teppp, it must be converted to a .teppp file by the user's scripts. In the case of a .read_data (a common file type for simulations in LAMMPS [36]), the convertor utility included with the source code can be used to convert it into a .teppp format. For example, if the file is in a read_data format, named systemA.read_data, it can be converted to .teppp format, as follows:

```
./convertor "/path/systemA.read_data"
```

This creates a new folder named "converted" within TEPPP (if it does not exist already) and a file systemA.teppp inside TEPPP/converted.

Table 1 shows the commands included in TEPPP. The table shows the commands for serial execution. For execution in parallel, an "_mpi" extension for each command is needed. The syntax for calling most of the commands is the same, and follows the convention

```
./command "/path/FILENAME" system specifications
```

where FILENAME is the name of the data file (including the path if not in the same directory) and the system specifications are (i) the chain length, (ii) the number of chains, (iii) the chain architecture (ring/linear), denoted by 0/1, respectively, (iv) (only for jones) the number of projections used in the computation of the Jones polynomial (it should be equal to 1, if ring architecture) and (v) periodic box dimensions (if in PBC). For the "_scan" commands, the additional system specifications (a) the initial scanning length, (b) the final scanning length and (c) the increments of scanning length must be added before the periodic box dimensions (if in PBC).

For example, to compute the writhe of all the chains in a system of 20 chains of 100 monomers each in a periodic cubic box of edge length 13.35315, which is already in .teppp format, called "systemA.teppp", one should type the following in the command line, from within the "build" directory of TEPPP, after compiling it:

```
./wr "../path/systemA.teppp" 20 100 1 13.35315
```

This will create an output directory, if it does not already exist, within the build directory, where a file "wr_out.txt" can be found. "wr_out.txt" has the writhe of each chain in the system, in the order in which the chains are read in the input .teppp file.

Similarly,

```
./lk "../path/systemA.teppp" 20 100 1 13.35315
```

will create a file "lk_out.txt" in the output directory which has the linking number of each pair of chains in the system. Since the linking number is symmetric, we compute the linking number $lk(i, j)$ of chains $i = 1, \ldots, n-1$ with $j = i+1, \ldots, n$, printed in that order.

The command

```
./jones "../path/systemA.teppp" 20 100 1 100 13.35315
```

will create a "jones_out.txt" file in the output directory that contains the Jones polynomial of each chain in the system computed using 100 different random projections.

The command

```
./periodic_wr "../path/systemA.teppp" 20 100 1 13.35315
```

will create a file "periodic_wr_out.txt" which has the periodic writhe (sum of linking numbers with periodic images) of each chain in the system, in the order in which the chains are read in the input .teppp file.

The command

```
./periodic_lk "../path/systemA.teppp" 20 100 1 13.35315
```

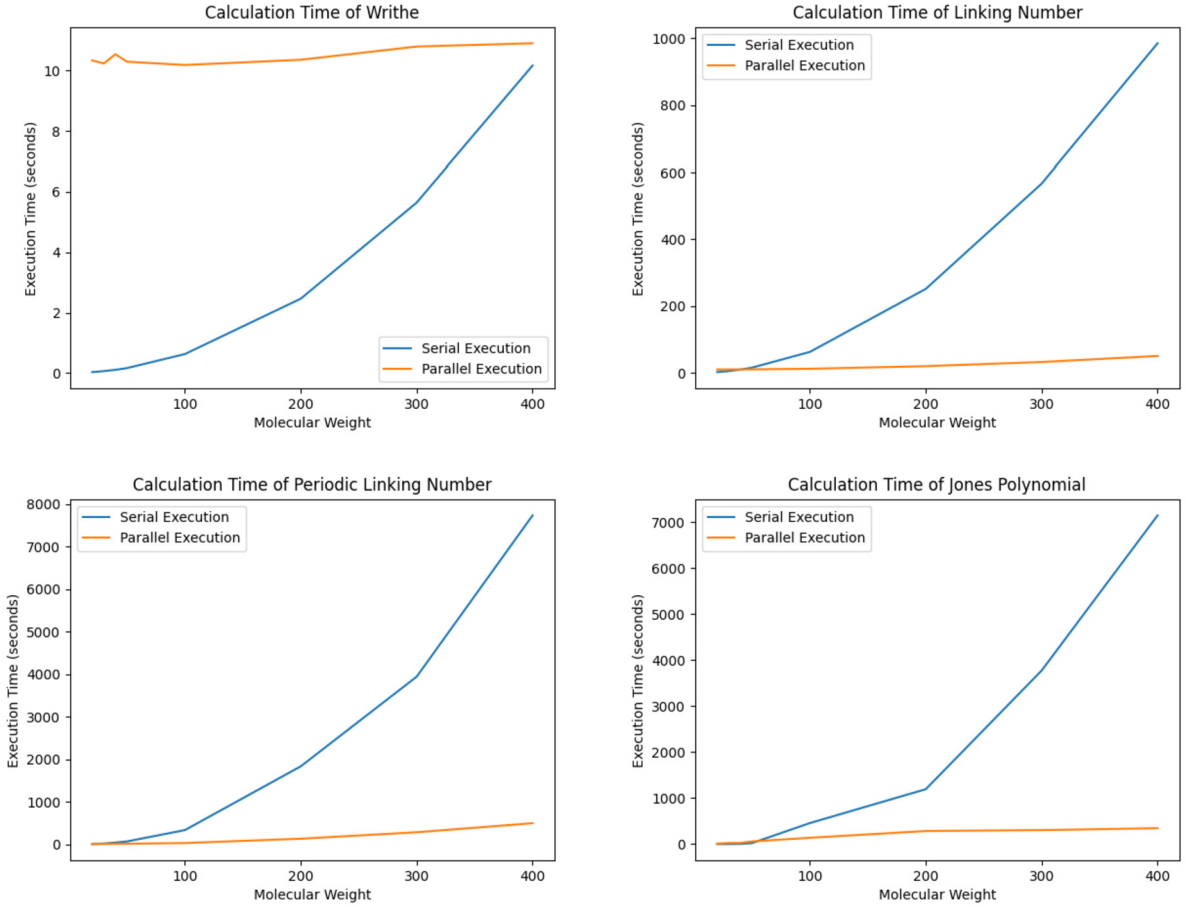


Fig. 4. Computation time for the writhe, the Gauss linking integral, the periodic linking number and the Jones polynomial in serial and in parallel for all chains or pairs of chains in a system of 100 chains of varying molecular weight at constant dimensionless number density 0.84.

will create a file "periodic_lk_out.txt" which has the periodic linking number of each pair of chains in the system. Since the periodic linking number is symmetric, we compute the periodic linking number $lk_P(i, j)$ of chains $i = 1, \ldots, n-1$ with $j = i+1, \ldots, n$, printed in that order.

The command

```
./wr_scan "../path/systemA.teppp" 20 100 5 10 5 13.35315
```

will create file "wr_scan_out.txt" that has the writhe of each segment of specified length of each chain in the system, in the order in which the chains are read in the input .teppp file.

Similarly,

```
./lk_scan "../path/systemA.teppp" 20 100 5 10 5 13.35315
```

will create file "lk_scan_out.txt" that has the linking number of each pair of segments of specified length between each pair of chains in the system, in the order in which the chains are read in the input .teppp file.

The command

```
./jones_scan "../path/systemA.teppp" 20 100 100 5 10 5 13.35315
```

will create file "jones_scan_out.txt" that has the Jones polynomial of each segment of specified length of each chain in the system, in the order in which the chains are read in the input .teppp file.

All the commands can be called in parallel. For example,

```
mpirun n -4 ./wr_mpi "../path/systemA.teppp" 20 100 1 13.35315
```

will create a directory wr_mpi inside the output directory, which contains the files of the parallel computation of the writhe of each chain in the system. The combination of the files should be the same as the output file wr_out.txt from the serial execution.

The computation time for the writhe, linking number, periodic linking number and Jones polynomial for polymer melts of 100 chains of varying length with a dimensionless number density 0.84 is shown in Fig. 4.

TEPPP contains the folders "main" and "include". The folder "main" contains the files wr.cpp, wr_scan.cpp, periodic_wr.cpp, lk.cpp, lk_scan.cpp, periodic_lk_scan.cpp, jones.cpp, jones_scan.cpp, which compute the corresponding topological measures. The folder "includes" contains files for elementary mathematical operations, as well as the file "funcs.h" which contains functions particular to TEPPP, such as unfolding wrapped coordinates in PBC, projecting a chain in 3-space onto a plane, generating a random vector on the 2-sphere, computing crossings in a projection, etc.

The computations of all the functions, except the Jones polynomial are computed without any counting of crossings and any projections, using the method described in [4]. The computation of the Jones polynomial however requires the study of projections of the chains and their crossings and for linear chains, it requires averaging over multiple projections. To do this, TEPPP stores the polynomials as an array of coefficients centered at t^0. The smoothings that are needed to compute the Jones polynomial in a projection direction are implemented

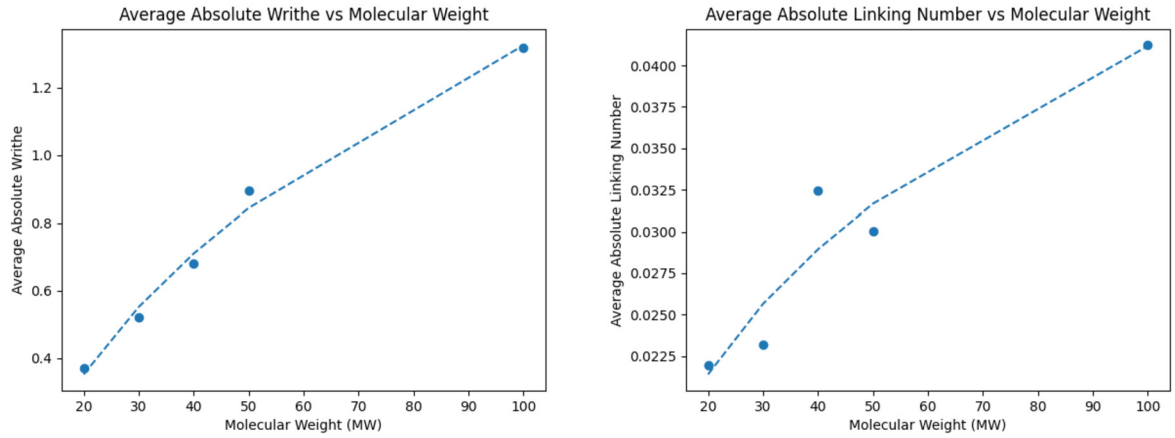


Fig. 5. The mean absolute writhe (Left) and periodic linking number (Right) in a polymer melt as a function of molecular weight. We see that self and pairwise entanglement increases with the length of the chains.

as a rearrangement of a connectivity list of each vertex (atom). When one edge has multiple crossings, the edge is split in multiple edges by introducing new vertices and updated the connectivity list so that each edge has only one crossing. More information about the computation of the topological measures can be found in Appendix A. Reidemeister I and II moves are employed by altering the crossings list when the corresponding conditions are met.

2.2.1. Usage of TEPPP for other polymer architectures

TEPPP can be used to measure entanglement of chains of varying architecture as follows:

For a star polymer, its arms can be seen as linear chains whose entanglement can be studied using TEPPP as explained for linear chains above. Special care is only needed for the linking number calculations, since for arms with the same root, the initial vertex (the root) should be omitted.

Similarly, for branched polymers, the backbones and the arms can be seen as linear chains, for which TEPPP can be applied to measure their entanglement. Attention is needed in this case when computing the linking of arms with the backbone they belong to, in order to omit their root from the arms coordinates.

In all these cases however, the chains that are included in one file must have the same molecular weight. In the same way, TEPPP can be used to study entanglement of dendric polymers in general.

3. Applications of TEPPP

In this section we present results obtained using the TEPPP to conformations of polymer melts and proteins.

3.1. Entanglement in polymer melts as a function of molecular weight

We use TEPPP to analyze the topology of linear polymer melts, as a function of Molecular Weight. Polymer melts of linear chains were simulated interacting via a repulsive Lennard-Jones (LJ) potential by molecular dynamics (MD), as a classical multibead FENE chain system with a dimensionless number density 0.84 at temperature $T = 1$, with a time step $\Delta t = 0.005$ within a velocity Verlet algorithm with temperature control. All samples were pre-equilibrated. The MD simulation was run a time span of the order $2\tau_R$ [32].

Fig. 5 shows the mean absolute Writhe and the mean absolute Periodic Linking Number as a function of molecular weight in a melt, computed using the wr and periodic_lk commands TEPPP. We see that both the self-entanglement and the pairwise entanglement increases with molecular weight.

The Jones polynomial was computed for the linear chains using the command “jones” in TEPPP. Fig. 6 shows the absolute value of the coefficients of the powers of t in the Jones polynomial as a function of molecular weight. The powers of t in the Jones polynomial and their coefficients show that knotting complexity increases at $N > 20$, and open trefoil knots appear at $N > 50$, near the entanglement length of these systems [32]. This work highlights how topology is related to characteristic length scales in polymer melts [38].

3.2. Entanglement in diblock polymers

The effects of polymer chain topology on the order-disorder transition of symmetric AB diblock copolymer melts were studied using coarse-grained molecular dynamics simulations in [21]. Specifically, chain conformations near the lamellar disordered transition in melts of symmetric (i.e., $50 - 50$) AB diblock copolymers of linear chains, rings, and trefoil knots at the same chain lengths were compared. TEPPP was used to identify the location of the knot in trefoil knotted diblock polymers, using the command “jones_scan”. The results showed that the knots became stretched and tightened in disordered melts close to the lamellar-disorder transition (see Fig. 7). This work highlights chain topology as an important factor in affecting microphase separation in block copolymers.

3.3. Topological analysis of the SARS-CoV-2 Spike protein

The SARS-CoV-2 Spike protein is a folded protein trimer which significantly re-arranges its conformation in 3-space to promote viral fusion. Its conformation is therefore very important for protein function [3]. We used TEPPP to study the local topological entanglement

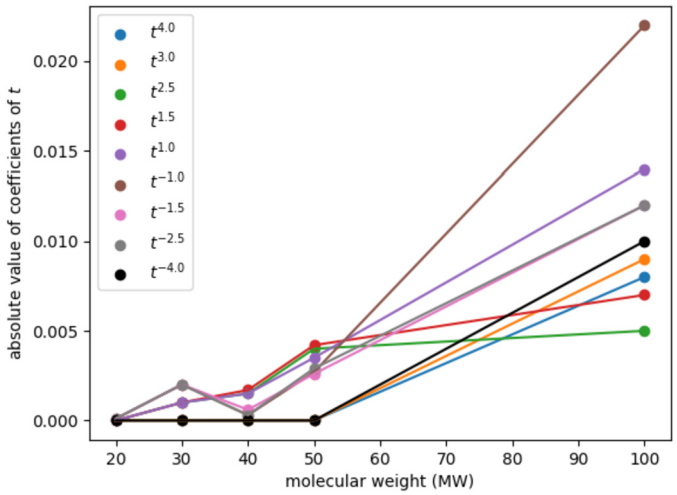


Fig. 6. The average coefficients of the powers in the Jones polynomial of linear chains in a melt as a function of molecular weight. We notice entanglement increasing at $N > 20$, in accordance with the onset of entanglements in those systems, as well as transitions at $N = 40$ and at $N \approx 70$ which correspond to critical lengthscales related to the entanglement length in the tube model of entangled polymers.

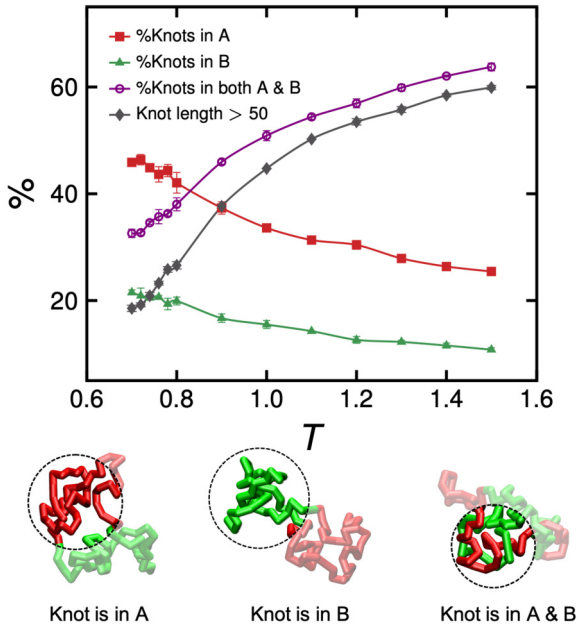


Fig. 7. The trefoil location and tightness in overall trefoil knotted diblock polymers is related to the lamella/disorder transition [21].
Figure from [21].

in the SARS-CoV-2 Spike protein in its closed pre-fusion state. By using the CA atoms of the protein structure 6zge obtained from the Protein Data Bank [7], we analyze the topology of the resulting chain. To do this, we used a pre-processing script that extracted the CA atoms of chain A and saved their coordinates in .teppp format. We use the "jones_scan" command to examine how the Jones polynomial varies along the protein backbone. We scan the protein at 100 CA atoms at a time, with a step size of 10. The highest power of the Jones polynomial that appears is $t^{\pm 5/2}$. The results on the coefficient of $t^{\pm 5/2}$ along the protein backbone are shown in Fig. 8. We see that the coefficient of $t^{5/2}$ is the largest in absolute value at 600-700 CA atoms, a region between the S1 and the S2 domains. This may reflect a global turn and packing of this region (similar to that shown in Fig. 2) that is important for the interaction between the S1 and S2 domains. It is interesting to point out that this region contains a cleavage site which is important for protein function. Our results thus point to a correlation between topological structure and protein function in this region, which may be used to understand protein folding.

4. Conclusions

In this paper we introduced the Topological Entanglement in Polymers and Proteins and Periodic systems (TEPPP) package. This is a software that can be used to analyze the topology of filamentous material. With this software one can compute the Writhe, the Gauss linking integral and the Jones polynomial of both open and closed curves in 3-space. Currently, this is the only software that computes the Jones polynomial without any closure scheme or 2-dimensional approximation scheme. Moreover, this is the only software that enables the calculation of the Periodic Writhe and Periodic linking number in systems employing Periodic Boundary Conditions, systems that occurs

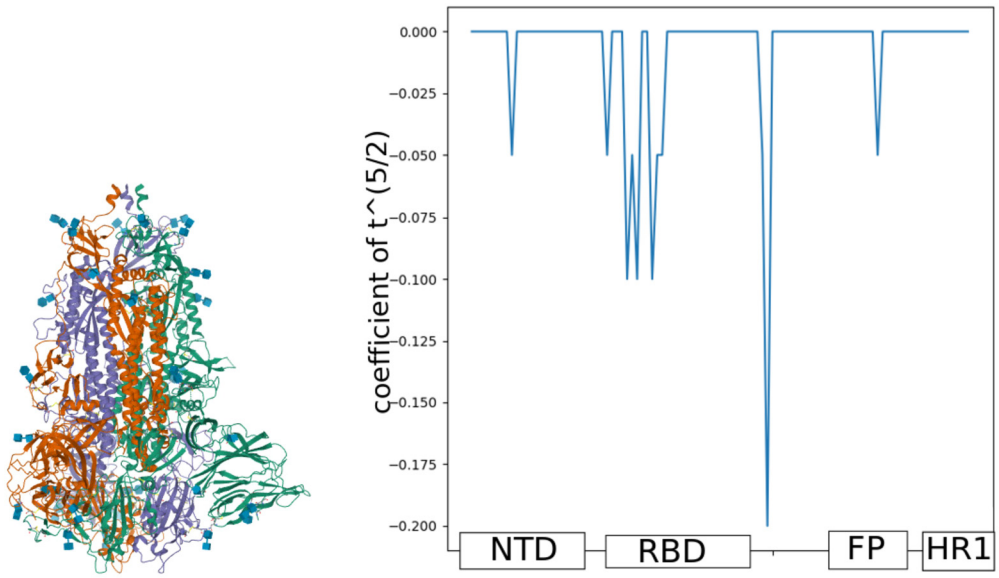


Fig. 8. The coefficient of $t^{5/2}$ of the Jones polynomial along the backbone of the SARS-CoV-2 Spike protein (PDB code 6zge). The backbone is analyzed in a window sliding approach of 100 CA atoms at a time. We see that the coefficient of $t^{5/2}$, which corresponds to non-trivial conformations, is the largest in absolute value at 600-700 CA atoms, a region between the S1 and S2 domain, which contains an important cleavage site.

in almost all molecular simulations of polymer melts and solutions. This software also enables the computation of all these parameters globally or locally along the filaments to identify interesting segments.

Our results on simulations of polymers showed that TEPPP can provide a rigorous definition of polymer entanglement that can lead to understanding important phase transitions or viscoelastic properties of polymer material. Our results on analyzing proteins from the PDB show that TEPPP can be used to detect conformational characteristics of proteins that are impossible to detect otherwise and which may be related to protein function. The software introduced in this manuscript is available online and can be easily used. By making topological tools available to polymer physicists, biologists and engineers, rigorous mathematical parameters can be computed to advance knowledge in materials research and biology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

TH and EP acknowledge support from NSF DMS #1913180, NSF #1925603 and NSF CAREER #2047587.

Appendix A. Topological and geometrical parameters in TEPPP

In this section we give the formal definitions of the topological parameters computed by TEPPP and we discuss the properties of these for both open and closed curves in 3-space.

A.1. The Gauss linking integral of open and closed curves

The Gauss linking integral of two oriented open or closed curves in 3-space is given by a double integral over the two curves:

Definition A.1 *(Gauss Linking Number).* The Gauss *Linking Number* of two disjoint (closed or open) oriented curves l_1 and l_2, whose parametrizations are $\gamma_1(t), \gamma_2(s)$ respectively, is defined as the following double integral over l_1 and l_2 [17]:

$$L(l_1, l_2) = \frac{1}{4\pi} \int_{[0,1]} \int_{[0,1]} \frac{(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))}{||\gamma_1(t) - \gamma_2(s)||^3} dtds, \tag{A.1}$$

where $(\dot{\gamma}_1(t), \dot{\gamma}_2(s), \gamma_1(t) - \gamma_2(s))$ is the *scalar triple product* of $\dot{\gamma}_1(t), \dot{\gamma}_2(s)$ and $\gamma_1(t) - \gamma_2(s)$.

For closed curves, the Gauss linking integral (in this case also called simply "linking number") is equal to half the algebraic sum of inter-crossings in the projection of the two curves in any projection direction, it is an integer and a topological invariant of the link. For

open curves, the Gauss linking integral is equal to the average of half the algebraic sum of inter-crossings in the projection of the two curves over all possible projection directions. It is a real number and a continuous function of the curve coordinates. For both open and closed curves, it measures how many times one curve interwinds around the other.

A measure for the degree of intertwining of the chain around itself is the Writhe, which is defined by taking the Gauss linking integral over one curve (instead of two).

Definition A.2 *(Writhe).* For a curve ℓ with arc-length parameterization $\gamma(t)$ is the double integral over l:

$$Wr(l) = \frac{1}{4\pi} \int_{[0,1]} \int_{[0,1]} \frac{(\dot{\gamma}(t), \dot{\gamma}(s), \gamma(t) - \gamma(s))}{||\gamma(t) - \gamma(s)||^3} dtds. \tag{A.2}$$

The Writhe is a continuous function of the chain coordinates for both open and closed chains.

For polygonal curves, the Gauss linking integral and the Writhe can be expressed as a sum of linking numbers over all inter or intra-, respectively, pairs of edges. For the Gauss linking integral between a pair of edges, one can avoid integration, by, equivalently, using a closed formula of the integral derived in [4]. This simple formula is also described in [6]. TEPPP employs this algorithm which enables the computation of the Gauss linking integral and the Writhe without the use of parametrizations and integration, simply by using the atomic coordinates, cross, dot products of the vectors they create and trigonometric functions.

We note that the Gauss linking integral can be applied in portions of any original curve. For example, we can apply the Writhe in sections of a single filament in order to examine how entangled each segment is, and where entanglement is stronger. Also, we can apply the Gauss linking integral in segments in pairs of chains to identify where the linking is higher.

The Gauss linking integral and the Writhe have been applied to many systems and shown it correlates with material properties [35,32,31,33,34,2].

A.2. The Jones polynomial

The Jones polynomial of an open or closed curve in 3-space can be defined using the normalized bracket polynomial:

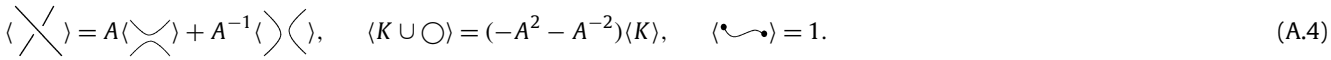
Definition A.3. Let l denote a curve in 3-space. Let $(l)_{\vec{\xi}}$ denote the projection of l on a plane with normal vector $\vec{\xi}$.

The normalized bracket polynomial of l is defined as:

$$f_{K(l)} = \frac{1}{4\pi} \int_{\vec{\xi} \in S^2} (-A^3)^{-wr((l)_{\vec{\xi}})} \langle (l)_{\vec{\xi}} \rangle dS \tag{A.3}$$

where the integral is over all vectors in S^2 except a set of measure zero (corresponding to irregular projections).

The following initial conditions and diagrammatic equations are sufficient for the skein computation of the bracket polynomial of projections of open curves (classical knotoids):



(A.4)

Note that for closed curves, the last diagrammatic equation consists of a circle, rather than an arc. For closed curves, the Jones polynomial defined in Eq. (A.3) is a topological invariant and coincides with the classical Jones polynomial of a knot [30]. In the case of a closed chain, the integral can be omitted, and the calculation can be obtained through the calculation of the integrand for any single projection of the curve in a plane. For open curves, the Jones polynomial is the average of the diagrammatic Jones polynomial of each projection over all possible projection directions, it has real coefficients and is a continuous function of the curve coordinates [30].

Recently, the Jones polynomial was rigorously defined for collections of linear chains [5]. This will enable the measurement of the multi-chain higher order entanglement in such systems.

A.3. Periodic linking number

Periodic entanglement occurs in many physical systems, such as textile weaves or simulations of polymer melts and solutions.

In a periodic system, each chain can link with multiple translations (images) of the other chains. The periodic linking number accounts all the linking that may occur from an image of one chain and all the translations of the other with no repetitions.

To define the periodic linking number it is necessary to give some definitions. First, we call the arcs inside the simulation cell, whose translations form one connected component (one chain), a generating chain. All the translations of the generating chain in the periodic system (infinite number), create an infinite number of translations of one connected component, copies of one curve in 3-space. We call free chain all the translations of that curve.

The periodic linking number is a convergent infinite sum for open curves. In practice, we use the local periodic linking number and the local periodic writhe, which are appropriate cut-offs of the periodic linking number and the periodic writhe, respectively [29,35].

Definition A.4. Let I, J be two free chains in a system with PBC. The local periodic linking number between I and J is defined as

$$LK_P(I, J) = \sum_{i \in I} Lk(I_0, J_i) \tag{A.5}$$

where I denotes the set of translations of J_0 that intersects the cells in which I_0 lies in.

The local periodic linking number is an invariant for closed curves and it is a continuous function of the chain coordinates for open curves [29].

In a periodic system, a chain may link with its own periodic images. In that case, this is accounted in the periodic Writhe:

Definition A.5. Let I be a free chain in a system with PBC. The local periodic writhe of I is defined as

$$Wr_P(I) = \sum_{i \in I} Lk(I_0, I_i) \tag{A.6}$$

where I denotes the set of translations of I_0 that intersect the cells in which I_0 lies in.

The periodic Writhe is a continuous function of the chain coordinates. For closed curves, the second term is a topological invariant.

References

[1] C.C. Adams, The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots, W.H. Freeman, New York, 1994.
[2] Q. Baldwin, E. Panagiotou, J. Theor. Biol. 529 (2021) 110854, https://doi.org/10.1016/j.jtbi.2021.110854.
[3] Q. Baldwin, B. Sumpter, E. Panagiotou, Polymers 14 (2022) 3014.
[4] T. Banchoff, Indiana Univ. Math. J. 25 (1976) 1171–1188.
[5] K. Barkataki, E. Panagiotou, Proc. R. Soc. A 478 (2022) 20220302.
[6] M.A. Berger, C. Prior, J. Phys. A 39 (2006) 8321–8348.
[7] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H.H. Weissig, I.N. Shindyalov, P.E. Bourne, Nucleic Acids Res. 28 (2000) 235–242.
[8] P. Dabrowski-Tumanski, M. Piejko, S. Niewieczerzal, A. Stasiak, J.I. Sulkowska, J. Phys. Chem. B 122 (2018) 11616–11625.
[9] P. Dabrowski-Tumanski, P. Rubach, D. Goundaroulis, J. Dorier, P. Sułkowski, K.C. Millett, E.J. Rawdon, A. Stasiak, J.I. Sulkowska, Nucleic Acids Res. 47 (2019) D367–D375.
[10] P. Dabrowski-Tumanski, P. Rubach, W. Niemyska, B.A. Gren, J.I. Sulkowska, Brief. Bioinform. 22 (2021) bbaa196.
[11] P-G. de Gennes, Macromolecules 17 (1984) 703–704.
[12] O. Delgado-Friedrichs, S. Hyde, M. O'Keeffe, O. Yaghi, Struct. Chem. 28 (2017) 39–44.
[13] F. Edwards, Proc. Phys. Soc. 91 (1967) 513–519.
[14] F. Edwards, J. Phys. A, Gen. Phys. 1 (1968) 15–28.
[15] M.E. Evans, V. Robins, S.T. Hyde, Acta Chryst. A69 (2013) 262–275.
[16] R. Everaers, S.K. Sukumaran, G.S. Grest, C. Svaneborg, A. Sivasubramanian, K. Kremer, Science 303 (2004) 823.
[17] K.F. Gauss, Kgl. Gesellsch. Wiss. Göttingen (1877).
[18] A.Y. Grosberg, A. Feigel, Y. Rabin, Phys. Rev. E 54 (1996) 6618.
[19] A.Y. Grosberg, S.K. Nechaev, E.I. Shakhnovich, J. Phys. Fr. 49 (1998) 2095–2100.
[20] A.Y. Grosberg, Y. Rabin, Phys. Rev. Lett. 99 (2007) 217801.
[21] T. Herschberg, J.M. Carrillo, B.G. Sumpter, E. Panagiotou, R. Kumar, Macromolecules (2021), https://doi.org/10.1021/acs.macromol.1c00780.
[22] M. Jamroz, W. Niemyska, E.J. Rawdon, A. Stasiak, K.C. Millett, P. Sulkowski, J. Sulkowska, Nucleic Acids Res. 43 (2015) D306–D314.
[23] V.F.R. Jones, Bull. Am. Math. Soc. 12 (1985) 103–112.
[24] V.F.R. Jones, Ann. Math. 126 (1987) 335–388.
[25] G. Kolesov, P. Virnau, M. Kardar, L. Mirny, Nucleic Acids Res. 35 (2007) W425–W428.
[26] M. Kröger, Phys. Rep. 390 (2004) 453–551.
[27] Y.L. La, C.C. Chen, J.K. Hwang, Nucleic Acids Res. 40 (2012) W228–W231.
[28] M.L. Mansfield, Nat. Struct. Biol. 1 (1994) 213–214.
[29] E. Panagiotou, J. Comput. Phys. 300 (2015) 533–573.
[30] E. Panagiotou, L. Kauffman, Proc. R. Soc. A 476 (2020) 20200124.
[31] E. Panagiotou, M. Kröger, Phys. Rev. E 90 (2014) 042602.
[32] E. Panagiotou, M. Kröger, K.C. Millett, Phys. Rev. E 88 (2013) 062604.
[33] E. Panagiotou, K.C. Millett, P.J. Atzberger, Polymers 11 (2019) 11030437.
[34] E. Panagiotou, K.W. Plaxco, in: Topology and Geometry of Biopolymers, in: AMS Contemporary Mathematics Series, vol. 746, 2020, pp. 223–233.
[35] E. Panagiotou, C. Tzoumanekas, S. Lambropoulou, K.C. Millett, D.N. Theodorou, Prog. Theor. Phys. Suppl. 191 (2011) 172–181.
[36] S. Plimpton, J. Comput. Phys. 117 (1995) 1–19.
[37] S. Redemann, J. Baumgart, N. Lindow, M. Shelley, E. Nazockdast, A. Kratz, S. Prohanska, J. Brugues, S. Fürthauer, T. Müller-Reichert, Nat. Commun. 10 (1038) (2017) 15288.
[38] M. Rubinstein, R. Colby, Polymer Physics, Oxford University Press, 2003.
[39] A. Stasiak, V. Katritch, J. Bednar, D. Michoud, J. Dubochet, Nature 384 (1996) 122.
[40] J.I. Sulkowska, E.J. Rawdon, K.C. Millett, J.N. Onuchic, A. Stasiak, Proc. Natl. Acad. Sci. 109 (2012) E1715.
[41] J.I. Sulkowska, P. Sulkowski, J. Onuchic, Proc. Natl. Acad. Sci. 106 (2009) 3119–3124.
[42] W.R. Taylor, Nature 406 (2000) 916–919.
[43] L. Tubiana, G. Polles, E. Orlandini, C. Micheletti, Eur. Phys. J. E 41 (2018) 72.
[44] C. Tzoumanekas, D.N. Theodorou, Macromolecules 39 (2006) 4592–4604.
[45] P. Virnau, L.A. Mirny, M. Kardar, PLoS Comput. Biol. 2 (2006) e122.
[46] P. Wadekar, P. Goel, C. Amanatides, G. Dion, R.D. Kamien, D.E. Breen, J. Eng. Fibers Fabr. 15 (2020) 1–15.