Published online 14 June 2018 *Nucleic Acids Research, 2018, Vol. 46, Web Server issue* ***W17–W24***
doi: 10.1093/nar/gky511

KnotGenome: a server to analyze entanglements of chromosomes

Joanna I. Sulkowska[1,2,*], Szymon Niewieczerzal[1], Aleksandra I. Jarmolinska[1,3], Jonathan T. Siebert[4], Peter Virnau[4] and Wanda Niemyska[1,5]

[1]Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland, [2]Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warsaw, Poland, [3]College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Banacha 2c, 02-097 Warsaw, Poland, [4]Johannes Gutenberg-Universität Mainz, Institut für Physik, Staudingerweg 9, Mainz, 55128, Germany and [5]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

Received March 26, 2018; Revised May 05, 2018; Editorial Decision May 21, 2018; Accepted May 23, 2018

ABSTRACT

The KnotGenome server enables the topological analysis of chromosome model data using three-dimensional coordinate files of chromosomes as input. In particular, it detects prime and composite knots in single chromosomes, and links between chromosomes. The knotting complexity of the chromosome is presented in the form of a matrix diagram that reveals the knot type of the entire polynucleotide chain and of each of its subchains. Links are determined by means of the Gaussian linking integral and the HOMFLY-PT polynomial. Entangled chromosomes are presented graphically in an intuitive way. It is also possible to relax structure with short molecular dynamics runs before the analysis. KnotGenome is freely available at http://knotgenom.cent.uw.edu.pl/.

INTRODUCTION

Topology plays an ever-increasing role in modern life sciences since the discovery (1) and artificial creation of knots in DNA (2,3) and proteins (4). Nowadays, up to 6% of proteins are known to be knotted (5–9), slipknotted (6,10), linked (11) or to contain lassos (12), even though for a long time it was believed that proteins should not be entangled. Analysis of such structures is now facilitated by powerful servers (13–18). It is well-known that the three-dimensional (3D) structure of the genome plays a critical role in regulating gene expression. Recent developments have for the first time enabled the determination of 3D structures of individual chromosomes and genomes based on Hi-C chromosome conformation contact data, e.g. in G1 phase nuclei of single haploid mouse embryonic stem cells. While model structures from (19) are highly knotted (20), and linked (21), and knots were also observed in (22,23), other models based on Hi-C data predict only few knots (24). Even though the abundance of entanglements in chromosomes is still controversial, there is a clear need to check model structures for entanglements, in particular when higher resolution data becomes available in the near future. To that end, we present the KnotGenome server—the first server that detects and characterizes knots in single chromosomes, as well as links between chromosomes, see Figure 1. This server is optimized to handle input data, which by far exceeds typical chain lengths of proteins and thus cannot be studied by existing servers. Prime and composite knots are determined by the computation of knot polynomials (either Alexander or HOMFLY-PT), and links are additionally characterized by the Gaussian Linking Number (GLN) (25). To give some hint about stability of entangled structures the server also enables to relax model data with short molecular dynamics runs before the analysis. This data provides new reaction coordinates—descriptors, which are crucial to improve current data, to understand the geometry of chromosomes and their interactions and to identify stable entangled configurations. In general, understanding entanglement is also a valuable resource for investigating the spatial structure-and-function relationship of genomes, and their potential role in regulating gene expression.

MATERIALS AND METHODS

The server has different workflows for a single chromosome and the whole cell. A user can choose various methods to close chains, as well as the level of detail of the topological analysis ranging from simple knot determination to a complete topological fingerprint with different resolutions.

*To whom correspondence should be addressed. Tel: +48 22 55 43 675 (Ext 320); Fax: +48 22 822 02 11; Email: jsulkowska@chem.uw.edu.pl

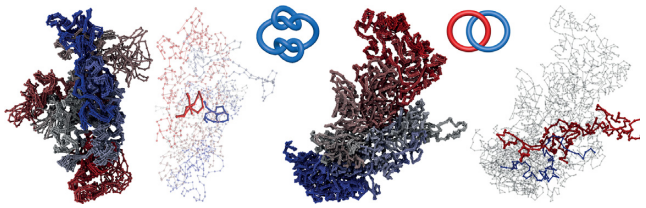

Figure 1. Left: Structure of chromosome 14 from cell 2 of (19) containing a $3_1\#3_1$ composite knot (marked by thick red and blue lines.). Right: visualization of a simple Hopf link between chromosomes 5 and 9 of cell 2. Again, the linked sections are marked by thick lines. Overlays of various models of the chromosomes are also displayed.

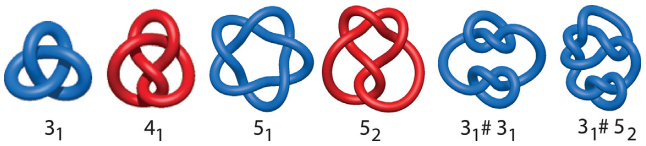

Figure 2. Schematical drawing of simple knots: trefoil (3_1), figure-eight (4_1), 5_1, 5_2 and composite knots $3_1\#3_1$ and $3_1\#5_2$.

The GLN method is used to spot locations of winding of pairs of chromosomes. The user can also relax model data via molecular dynamics simulations before analysis. Results are visualized via interactive matrices, pie charts and knots and links are displayed on the chromosomes. Output data provide detailed information for each knot, slipknot, and link, which can be downloaded for further analysis.

In the following we would like to provide a brief introduction to knots, links, GLN method, closures and optional relaxation procedures before describing the server interface in detail.

Knots

Mathematically, knots are only well-defined in closed curves. In order to investigate the topological state of a single chromosome, the endpoints need to be connected in a well-defined manner as detailed in the section on closures below and in the online help. Once a linear chromosome chain has been closed, knots can be categorized by the minimal number of crossings after a projection onto a plane. An unknotted loop is called the trivial knot, or unknot, and is denoted by 0_1. The simplest nontrivial knot is called trefoil (denoted also by 3_1). There is one knot with four crossings (4_1), there are two different knots with five crossings (5_1, 5_2), three with six crossings and from there on the number of different knots increases exponentially as a function of the crossing number. In the notation used above, the first number denotes the minimal number of crossings, while the subscript number is just an index to distinguish between knots with the same crossing number. Figure 2 shows schematically the first few non-trivial knots including so-called composite knots ($3_1\#3_1$ and $3_1\#5_2$), which consist of multiple knots embedded on the same chain. It is worth noting that knots (and composite knots) cannot be transformed into another knot type without breaking the chain. So once the chain is closed, the knot type embedded on the chain is fixed as well.

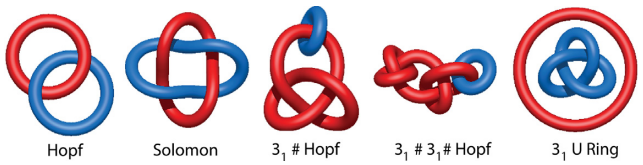

Figure 3. Examples of the notation for links.

Sizes and locations of knots in a chromosome can be obtained by successively clipping beads from both ends of a chain before determining the knot type for the remaining subchain. Results are shown in the form of a so-called knotting fingerprint matrix (6,10,13), which displays the topology of all subchains. An example of such a fingerprint matrix is shown in 'Discussion' section of data presentation. From a technical point of view, knots are determined with the HOMFLY-PT polynomial (26,27) using an implementation by Ewing and Millett (28) or alternatively the Alexander polynomial after applying the so-called KMT reduction (13,29,30). While the HOMFLY-PT polynomial is more powerful in its ability to distinguish between complicated knots and also provides information on chirality of knots, the Alexander polynomial is much faster from a computational point of view. For details the reader is referred to the online help section.

Links

While knots provide information on self-entanglements of single chromosomes, links describe entanglements between different chromosomes in a cell. Mathematically, links are a collection of circles or knots which may be linked together. As in the case of knots, links are only properly defined for closed curves and therefore the endpoints of the chromosomes involved need to be connected in a well-defined manner as detailed in the Closure section. The simplest link is the so-called unlink which consists of finitely many disjoint circles (unknots). The Hopf link is the simplest non-trivial link which consists of two circles linked together once (leftmost figure in Figure 3). In the context of our server, we use a more informal convention. Here, an unlink refers to disjoint knots (see rightmost figure in Figure 3) and a Hopf link refers to a 2-component link between two knots, which are linked once (see 3_1#Hopf and $3_1\#3_1$#Hopf in Figure 3).

GLN method

From a technical point of view, links are identified with HOMFLY-PT polynomials, as in (14). The GLN is used as another (computationally faster) measure of entanglement between a pair of chromosomes (25). The GLN indicates how many times (and in which direction) one chromosome winds around the other one. Moreover the GLN provides more information about the entanglements between chromosomes since it enables to identify the specific fragment of the chromosome winding around another one. More information on the classification and determination of links can be found online.

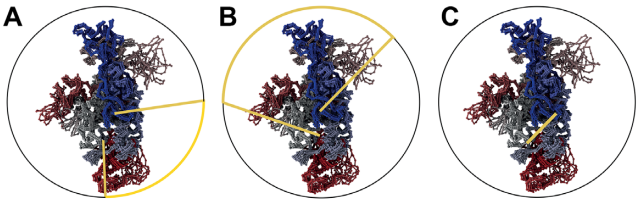

Figure 4. Methods used in KnotGenome to connect chromosome endpoints before link analysis. (**A**) Random closure method. (**B**) The center of mass method. (**C**) Direct closure method.

Closure

As indicated in the previous sections, the determination of knots and links requires that the endpoints of the respective chromosomes under consideration are connected first in a well-defined manner. Making this choice optimally is the first difficulty that has to be overcome when analyzing chromosomes—the closure may create knots on its own and has a large influence on the computational effort involved.

This server essentially provides three options to close the chain: (i) the random closure method (Figure 4A), where chromosome endpoints are connected to two points on a large sphere chosen randomly a certain number of times. Each time, these two points are connected by an arc lying on the surface of the sphere. The most frequently observed knot type is then associated with that chain as its dominant knot type. As a default option, 10 closures are calculated. The user can increase the number of random closures to 100 to improve the quality of the data. (ii) The center of mass method (Figure 4B). The endpoints are connected to two points on the sphere along the direction of a connection line between the center of mass and the respective endpoints. (iii) The direct closure method (Figure 4C), which connects chromosome endpoints by the shortest interval. Note that the statistical closure method (option i) minimizes the chances to 'detect' a knot, which was artificially created by the closure and is not really embedded in the structure as we average over different closures. Therefore, this method generally yields the best results although the computational effort increases linearly with the number of attempted closures. For a quick scan of the topology, we recommend option (ii).

Relaxation

Currently, the server cannot compute structures directly from a contact matrix. However, it can perform an optional relaxation via a short equilibrium molecular dynamics simulation of a given precomputed structure to estimate the robustness of the topology. This might not be advisable in some cases as the underlying model used to generate the 3D structure of the chromosomes is essentially replaced by our generic polymer model during relaxation. Nonetheless, in some cases the procedure may provide insights into the uncertainty of the topology estimation. An entanglement may be considered to be stable if it persists during relaxation.

Specifically, using the provided structures as starting configurations, the relaxation is performed using the GROMACS software (31) and a structure-based polymer model follwing method described here (21). The potential function has the following form:

$$V = \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_a(cos(\theta) - cos(\theta_0))^2 + \sum_{\text{dihedrals}} [k_d^1(1 + cos(\phi - \phi_0)) + k_d^3(1 + cos3(\phi - \phi_0))] + \sum_{\text{contacts}} 4\alpha\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \sum_{\text{non-contacts}} 4\epsilon \left(\frac{\sigma}{r}\right)^{12},$$

with the following force constants: $k_b = 20000.0\varepsilon/\text{nm}^2$, $k_a = 20\varepsilon$, $k_d^1 = 1.0\varepsilon$, $k_d^3 = 0.5\varepsilon$, $\alpha = 0.2$. Non-bonded interactions are introduced for pairs of beads within and between all chains. Pairs that are separated by a distance not larger than $r_{\text{cutoff}} = 2.0$nm, and not shorter than r_{min} in the starting structure are considered to be in contact. Pairs of non-bonded beads which are closer than r_{min} in the starting structure do not interact with each other (remain transparent to each other). All particles with a starting distance larger than r_{cutoff} only interact repulsively (non-contacts). To avoid excessive overlap for bonded interactions, each bead which is within a distance of $r_{\text{min}} = 0.6$nm of the preceding bead along the chain, is removed. By default, the system is propagated for 200 000 time steps with the time step $dt = 0.0005$ at temperature $T = 120$. For these settings, fluctuations of the whole chromosome cell lead to a RMSD of about $0.7 - 1.0$nm.

A detailed discussion of rules for stability including examples for links and knots are presented on the KnotGenome website.

Computational costs

To provide rough estimates of the computational costs associated with the procedures outlined above, average waiting times to get results on the server are summarized in Table 1. These benchmarks were done using the structures provided in ref. (19). They include determining knot types and knotting fingerprints for single chains. Links and the GLN were tested for both pairs of chromosomes as well as whole cells. Both available closure types (center of mass closure and random closure based on 10 closure attempts) are evaluated for all those cases respectively. As expected, the random closure is computationally much more expensive. Also computations of a full knotting fingerprint of a chromosome or links and the GLN for a full cell are computationally much more expensive than determination of the main knot type and linking and the GLN of pairs of chromosomes. Therefore, starting with these simpler analyses allows to gain quick insights into the topology of the analyzed structures. Nonetheless, all computations are reasonably fast. Even applying the random closure, determination of the full knotting fingerprint or finding links and the GLN for a full cell take less than an hour on average.

Server technicalities

Chromosomes are visualized with JSmol (HTML5/JavaScript version) with all available functions. The whole cell matrices, GLN matrices and pie

Table 1. Average computational costs for determining entanglement types in single chromosomes or whole cells as well as costs for constructing the chromosome fingerprint matrix (average time needed to get results on the server without prior relaxation)

Average time to compute and display results on the webpage	Center of mass closure [minutes]	Random closure (10 closure attempts) [min]
Knotting fingerprint	5	45 ± 32
Main knot type	1	1 ± 1
Link and GLN value between two chromosomes	1	1 ± 1
Link and GLN values for the whole cell	6	29 ± 18

The first three times are determined based on cell no. 2 from Ref. (19)—20 chromosomes of the length between 500 up to 2000 (beads), 190 chromosome pairs. The average time to determine entanglement of the whole cell is determined based on eight cells from Ref. (19).

charts are generated using Plotly. Algorithms to determine knots, links, and GLN method were implemented using the C++ language. The website is written in Python with the Flask framework dynamically generating HTML pages using Apache2 with WSGI. The whole service is installed on multicore Linux nodes.

SERVER INTERFACE

Server input and output files

Users can upload and analyze two types of data: either a single chromosome, or a whole cell. A whole cell includes between 2 and 40 chromosomes in a single input file, and is also suitable to analyze different conformations of a single chromosome e.g. a set of frames from a trajectory. The input data required to analyze topology of chromosome chains consists of a file with atomistic coordinates, in PDB or XYZ format (each monomer corresponding to, e.g. 100 000 bp). Sample input files for each options are provided on the website.

The results are presented graphically as discussed below. Additionally, a user can download a converted input file, the results of the analysis (the type of the topology with its probability and information arising from GLN analysis) as a text file for further analysis.

Users can name their jobs, hide them on the job queue and request to be notified by e-mail when their job finishes. E-mail addresses are stored only until the job finishes, all other data about the jobs is stored for 2 weeks.

Advanced options

The server was optimized to give the best results possible in reasonable time. By default the matrix fingerprint is calculated every seventh bead (matrix resolution), but this value can be changed. This reduction should not change the type of the fingerprint while significantly reducing the computing time. Matrices demonstrating different resolutions and the corresponding computation times are presented on the web page. Moreover the user can change the knot strength threshold for which only Alexander polynomial is calculated, which is much faster than HOMFLY-PT.

Online documentation

The server is supported by extensive, well-organized online documentation which includes: descriptions of the knot and link detection and classification methods, the GLN method and the relaxation technique, a list of different topological types identified in the tested genome data and instructions on how to use the server efficiently and how to interpret the results. Examples of input data, as well as the corresponding results are available.

DISCUSSION OF DATA PRESENTATION

The fundamental information about each submitted job—topology of chromosome(s)—is presented in the same way, consisting of: job status, chromosome structure presentation(s), a visualization of the closure method, knot and link probability pie charts and a table with the identified list of knots or links and their respective probabilities, as well as information about their robustness if chosen by the user. Moreover: (i) for a single chromosome a topological fingerprint can be determined; (ii) for a pair of chromosomes, the magnitude and location of winding is determined; (iii) for more than one pair of chromosomes additionally an overview of the identified topology is presented with interactive matrices.

At the top of each result page, the first tab indicates the status of a submitted job and summarizes options chosen by a user. The most important information is the success of calculation and the results. The user is informed whether a knot or link has been identified. Additionally, information about the method used to close the chromosome chain is included. If the random closure method was chosen, the number of random closure points is shown. In the following, we will discuss a few sample jobs to showcase how to interpret the data output.

Single chromosome: the whole chain

Figure 5 shows a sample output from a request to analyze a knot in a single chromosome using the random closure method (100 closures). On the left hand side of the main panel the chromosome structure is visualized. The user can use all of the standard JSmol capabilities, including zoom or changing the colors. On the right hand side, the various knot types that have been detected are visualized via the pie chart. The majority knot type (in this case a 5_2 knot with negative chirality as indicated by the minus sign) is displayed inside the pie chart, which also visualizes the probability of observing a particular knot type. Upon pointing the cursor at any part of the pie chart the corresponding knot type picture is presented inside the chart along with its likelihood. The same information is also displayed in a table below.

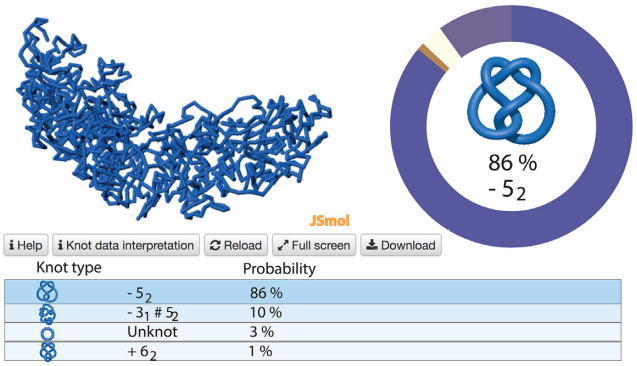

Figure 5. Typical data output from a request to analyze the knot type in a single chromosome using the random closure method with 100 closures.

Single chromosome: the topological fingerprint

Figure 6 shows a sample output displaying a so-called knotting fingerprint matrix. This matrix displays the topology of all knotted subchains of a chromosome and is therefore expensive from a computational point of view. The server offers different levels of graining to speed up computation (for details see the web page). The matrix diagram is constructed as follows. For a given chain of length N, all of its subchains spanning between residues k and l (for $1 \leq k \leq l \leq N$) are analyzed. If a subchain from k to l is knotted, then a point with coordinates (k, l) is denoted by a color corresponding to a particular knot. The intensity of color represents the probability of identifying a given knot (if the random closure method is applied). In this particular example the chain contains a slipknot, i.e. the chain as a whole is unknotted, but contains a knotted subchain.

For example, if one wants to know if the section between monomer 239 and 560 contains a knot we can look at position $x = 239$ and $y = 560$ and check if the point is colored or not. In this particular example, the point is green, which indicates that the program has detected a trefoil knot for that section. On the other hand, if we check, for example the section between 239 and 585 (the end of the chain), no knot was detected and the point contains no color (indicating an unknot). This is possible because the last part of the chain (the slipknot loop and tail (32)) threads through a loop of the trefoil dissolving the knot in the process.

The 'knotting fingerprint of a chromosome' encodes information about locations of the 'knot core' (encoding the smallest segment required to observe a knot), 'knot tails', 'slipknot loops' and 'slipknot tails', as defined in ref. (6), for each identified knot. These values and the corresponding type of knot are listed below the matrix in the knot table.

The matrix diagram is interactive: after choosing a knot type (if more than one knot type is detected) from the table, the data corresponding to this knot is displayed in the diagram and in a graphical form directly on the chromosome. By default the data corresponding to the knot formed by the whole chain (for knotted chromosomes), or the most complicated slipknot (for chromosomes with slipknots) is shown in the diagram.

A user can download, for example a picture representing the matrix and the corresponding raw data in a simplified XYZ format for further analysis.

Pair of chromosomes: links analysis

Comprehensive information about entanglement of a pair of chromosomes is presented in three interactive matrices, as shown in Figure 7:

- Linking of chromosomes—diagonal elements of this interactive matrix describe knot types of individual chromosomes while lower off-diagonal elements characterize linking between chains. The respective color codes are displayed below the matrix. If the random closure method was applied, the most probable type of entanglement is presented.
- Whole GLN—an interactive matrix describing the winding between two whole chromosomes as determined with the GLN approach. The color scale is adjusted for each pair of chromosomes. When the whGLN is bigger than 1 or lower than −1 there is a high probability that the chromosomes are linked.
- *Max*|*GLN*|—an interactive matrix, which describes maximal local winding between fragments of two chromosomes. The color scale is the same to the scale used for whGLN. Note that *max*|*GLN*| is equal *max*(*maxGLN*, −*minGLN*), where *minGLN* and *maxGLN* denote respectively minimum and maximum values of GLN between fragments of two chromosomes. Also note that *max*|*GLN*| may indicate slipknot geometry.

Upon pointing the cursor on a matrix element detailed information about the topology, and GLN is provided. Upon clicking on an element of the matrix, a new web page is provided containing the graphical representation of the chromosomes and knot or link likelihood, and an extended version of the knot or link table is displayed. Figure 8 shows such new web page for a pair of links. All matrices can be, for example zoomed in/out, modified and saved using the open source tool Plotly, which allows interactive data visualization (https://plot.ly/). The topological and structural details about the most probable type of entanglement (when the probabilistic closure method is used) are stored in the Knot and Link table below the matrices. More details on the definition of GLNs are presented on the KnotGenome website.

Links analysis: the whole cell

The user can upload and analyze a whole cell. In this case, the entanglements are presented via interactive matrices. Their dimensions correspond to the number of chromosomes. Thus, all topological methods, their results and techniques of presentation used for a single pair of chromosome are also available for multi-pair systems. The overview of a typical output for 20 chromosomes is presented in Figure 9.

METHOD VALIDATION AND COMPARISON WITH OTHER DATABASES

First, algorithms to determine knots, knotting fingerprints, links and GLN were validated earlier on all proteins de-

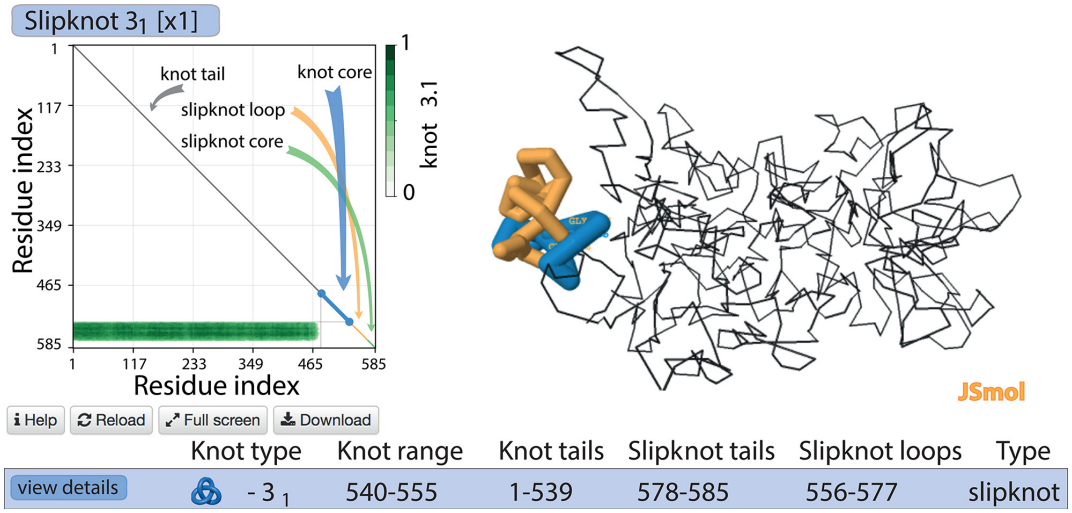

Figure 6. An example of data output for a slipknotted chromosome. The analysis reveals that the entire analyzed chain forms an unknot and has a subchain forming a 3_1 knot. Diagram in the top left: knotting fingerprint revealing the positions of the subchain forming a 3_1 knot. Top right: graphical representation of chromosome in JSmol, with the knotted core, slipknot and knot tails highlighted in various colors (knot core in blue, slipknot loop in orange and the knot tail in gray). In gray the knot tail is shown. Table at the bottom: detailed data about knots and slipknots formed by backbone subchains, based on 100 closures.

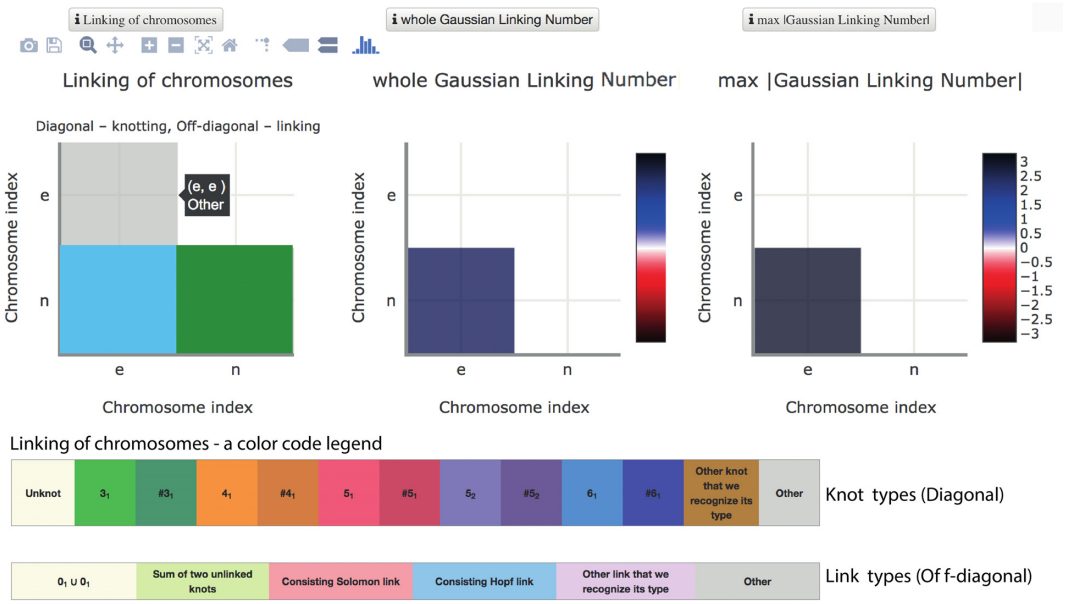

Figure 7. Sample output for analysis of links between two chromosomes. Left: interactive linking matrix. Diagonal elements describe knot types of individual chromosomes while lower off-diagonal elements characterize linking between chains. The respective color codes are displayed below the matrix. Middle and right: interactive matrices, which display whole GLNs and max |GLN|.

posited in the PDB (13,14,17,25). Second, all methods that we used were developed further, optimized to take into account properties of chromosomes (20,21) and validated on 1600 single chromosomes and 1520 pairs of chromosomes (from eight cells) provided by ref. (19). These results, including all topological information, are also available in the 'Job' section of KnotGenome. The analysis provided by KnotGenome cannot be performed by any other available server that analyzes protein structures and topology.

To our knowledge, KnotGenome is the first server able to identify prime and composite knots in single chromosomes and links between chromosomes. In comparison to other servers of knotted or linked structures, such as KnotProt (http://knotprot.cent.uw.edu.pl/), and LinkProt (http://linkprot.cent.uw.edu.pl), this server: (i) is able to analyze polymers composed of up to 5000 beads, (ii) identifies composite knots, (iii) offers four different types of closures, (iv) uses the GLN to identify the entangled regions and (v) estimates the robustness of the links via relaxation.

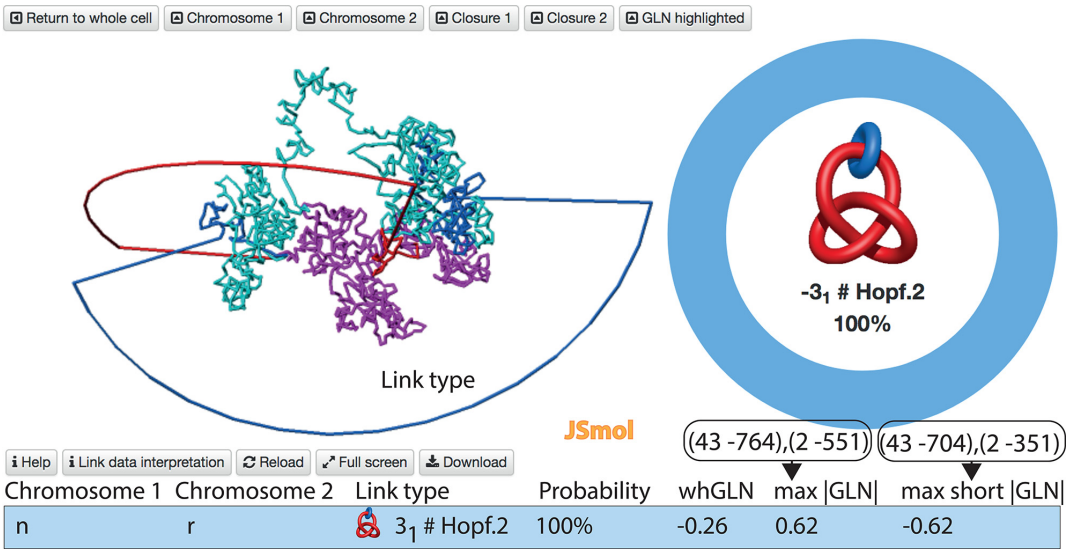

Figure 8. Sample output for analysis of links between two chromosomes. The cyan and magenta indicate the shortest fragments of chromosomes which wind around each other based on max*short*|*GLN*|. The exact location can be seen by clicking on GLN highlighted.

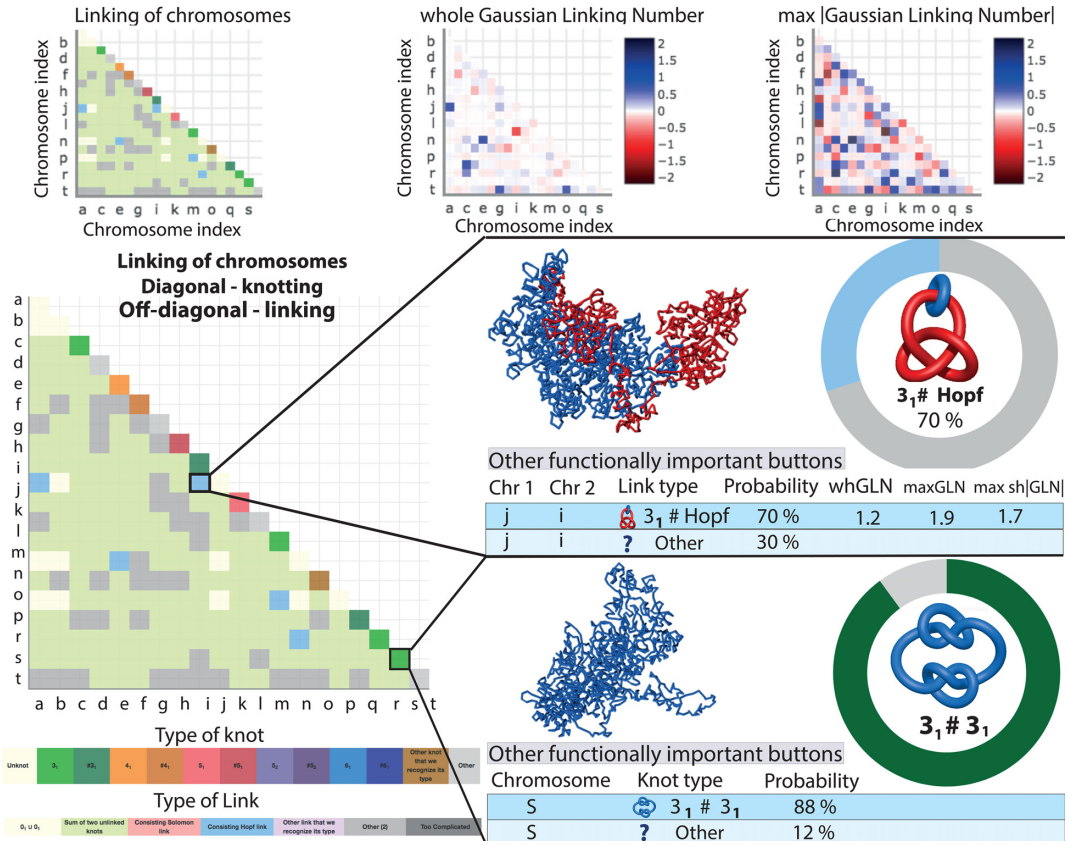

Figure 9. Overview of a comprehensive topological analysis of a whole cell, e.g. 20 pairs of chromosomes. Top row is the miniature figure of interactive matrices: linking chromosomes (showing with color code identified knots, links), whole GLN and max |*GLN*| which shows entanglement computed based on Gaussian linking integral, which is another measure of entanglement between a pair of chromosomes. Upon pointing the cursor on the element of matrix detail information about topology is provided. Upon clicking on the element of matrix a new subpage will be provided containing the Graphical representation of the chromosomes and knots or links likelihood, and the more extended version of the knot or link table is displayed.

SUMMARY

With KnotGenome we present a unique and highly optimized server, which allows researchers to determine topological properties of model chromosome structures based on Hi-C chromosome conformation contact data or any other method and incorporate this information in their workflows. Our server requires 3D coordinate files as input and does not model structures based on contact matrices, but provides information on the topological state of chromosomes based on the structures provided by the user. The server allows non-specialists to determine knot types and locations of knots and slipknots in individual chromosomes, as well as links between pairs of chromosomes. This structural information is crucial to improve current models, and to understand the geometry of chromosomes and their interactions. It will also help in further understanding of spatial structure-and-function relationships of genomes and help elucidate the role of entanglements in regulating gene expression at the chromosome level.

ACKNOWLEDGEMENTS

We would like to thank Pawel Pasznik, Jakub Ratajczak, and Tappest team, for their help with constructing the server.

FUNDING

National Science Centre [2012/07/E/NZ1/01900 to J.I.S.]; EMBO Installation Grant [2057 to J.I.S.]; DSM grant 501-D313-86-0117000-03 to WN. Funding for open access charge: EMBO YIP to JIS.
Conflict of interest statement. None declared.

REFERENCES

1. Macgregor,H. and Vlad,M. (1972)Interlocking and knotting of ring nucleoli in amphibian oocytes. *Chromosoma*, **39**, 205–214.
2. Liu,L.F., Liu,C.-C. and Alberts,B.M. (1980) Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break. *Cell*, **19**, 697–707.
3. Dean,F.B., Stasiak,A., Koller,T. and Cozzarelli,N.R. (1985) Duplex DNA knots produced by Escherichia coli topoisomerase I. Structure and requirements for formation. *J. Biol. Chem.*, **260**, 4975–4983.
4. King,N.P., Jacobitz,A.W., Sawaya,M.R., Goldschmidt,L. and Yeates,T.O. (2010) Structure and folding of a designed knotted protein. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 20732–20737.
5. Dabrowski-Tumanski,P. and Sulkowska,J.I. (2017) To tie or not to tie? That is the question. *Polymers*, **9**, 454.
6. Sułkowska,J.I., Rawdon,E.J., Millett,K.C., Onuchic,J.N. and Stasiak,A. (2012) Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Nat. Acad. Sci. U.S.A.*, **109**, E1715–E1723.
7. Virnau,P., Mallam,A. and Jackson,S. (2010) Structures and folding pathways of topologically knotted proteins. *J. Phys. Cond. Mat.*, **23**, 033101–033118.
8. Bolinger,D., Sulkowska,J.I., Hsu,H.-P., Mirny,L.A., Kardar,M., Onuchic,J.N. and Virnau,P. (2010) A Stevedore's protein knot. *PLoS Comput. Biol.*, **6**, e1000731.
9. Virnau,P., Mirny,L.A. and Kardar,M. (2006) Intricate knots in proteins: function and evolution. *PLoS Comput. Biol.*, **2**, e122.
10. King,N.P., Yeates,E.O. and Yeates,T.O. (2007) Identification of rare slipknots in proteins and their implications for stability and folding. *J. Mol. Biol.*, **373**, 153–166.
11. Dabrowski-Tumanski,P. and Sulkowska,J.I. (2017) Topological knots and links in proteins. *Proc. Natl. Acad. Sci. U.S.A.*, **114**, 3415–3420.
12. Niemyska,W.E.A. (2016) Complex lasso: new entangled motifs in proteins. *Sci. Rep.*, **6**, 36895.
13. Jamroz,M., Niemyska,W., Rawdon,E.J., Stasiak,A., Millett,K.C., Sułkowski,P. and Sulkowska,J.I. (2015) KnotProt: a database of proteins with knots and slipknots. *Nucleic Acids Res.*, **43**, D306–D314.
14. Dabrowski-Tumanski,P., Jarmolinska,A.I., Niemyska,W., Rawdon,E.J., Millett,K.C. and Sulkowska,J.I. (2017) LinkProt: a database collecting information about biological links. *Nucleic Acids Res.*, **45**, D243–D249.
15. Dabrowski-Tumanski,P., Niemyska,W., Pasznik,P. and Sulkowska,J.I. (2016) LassoProt: server to analyze biopolymers with lassos. *Nucleic Acids Res.*, **44**, W383–W389.
16. Kolesov,G., Virnau,P., Kardar,M. and Mirny,L.A. (2007) Protein knot server: detection of knots in protein structures. *Nucleic Acids Res.*, **35**, W425–W428.
17. Gierut,A.M., Niemyska,W., Dabrowski-Tumanski,P., Sułkowski,P. and Sulkowska,J.I. (2017) PyLasso: a PyMOL plugin to identify lassos. *Bioinformatics*, **33**, 3819–3821.
18. Gierut,A.M., Dabrowski-Tumanski,P., Niemyska,W., Millett,K.C. and Sulkowska,J.I. (2018) PyLink: a PyMOL plugin to identify links. *Under Rev.*, **1**, 1–2.
19. Stevens,T.J., Lando,D., Basu,S., Atkinson,L.P., Cao,Y., Lee,S.F., Leeb,M., Wohlfahrt,K.J., Boucher,W., O'Shaughnessy-Kirwan,A. *et al.* (2017) 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature*, **544**, 59–64.
20. Siebert,J.T., Kivel,A.N., Atkinson,L.P., Stevens,T.J., Laue,E.D. and Virnau,P. (2017) Are there knots in chromosomes? *Polymers*, **9**, 317.
21. Niewieczerzal,S., Niemyska,W. and Sulkowska,J.I. (2018) Are there links in chromosomes? *Under Rev.*, **1**, 1–10.
22. Di Pierro,M., Zhang,B., Aiden,E.L., Wolynes,P.G. and Onuchic,J.N. (2016) Transferable model for chromosome architecture. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 12168–12173.
23. Valdes,A., Segura,J., Dyson,S., Martinez-Garcia,B. and Roca,J. (2018) DNA knots occur in intracellular chromatin. *Nucleic Acids Res.*, **46**, 650–660.
24. Mirny,L.A. (2011) The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.*, **19**, 37–51.
25. Niemyska,W., Millett,K.C. and Sulkowska,J.I. (2018) The Gaussian linking number—new method for studying lassos in proteins. *Under Rev.*, **1**, 1–10.
26. Freyd,P., Yetter,D., Hoste,J., Lickorish,W.R., Millett,K. and Ocneanu,A. (1985) A new polynomial invariant of knots and links. *Bull. Am. Math. Soc.*, **12**, 239–246.
27. Przytycki,J.H. and Traczyk,P. (1988) Invariants of links of Conway type. *Kobe J. Math.*, **4**, 115–139.
28. Ewing,B. and Millett,K.C. (1991) A load balanced algorithm for the calculation of the polynomial knot and link invariants. *Mathematical Heritage of C. F. Gauss*, **1**, 225–266.
29. Koniaris,K. and Muthukumar,M. (1991) Self-entanglement in ring polymers. *J. Chem. Phys.*, **95**, 2873–2881.
30. Virnau,P., Kantor,Y. and Kardar,M. (2005) Knots in globule and coil phases of a model polyethylene. *J. Am. Chem. Soc.*, **127**, 15102–15106.
31. Hess,B., Kutzner,C., Van Der Spoel,D. and Lindahl,E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.
32. Sulkowska,J.I., Sułkowski,P. and Onuchic,J.N. (2009) Jamming proteins with slipknots and their free energy landscape. *Phys. Rev. Lett.*, **103**, 268103.