Bioinformatics, 34(19), 2018, 3402–3404
doi: 10.1093/bioinformatics/bty365
Advance Access Publication Date: 2 May 2018
Applications Note

OXFORD

Structural bioinformatics

Knoto-ID: a tool to study the entanglement of open protein chains using the concept of knotoids

Julien Dorier[1,2,*,†], Dimos Goundaroulis[2,3,†], Fabrizio Benedetti[1,2] and Andrzej Stasiak[2,3]

[1]Vital-IT, SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland, [2]Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland and [3]SIB Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

*To whom correspondence should be addressed.
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Associate Editor: Alfonso Valencia

Received on November 9, 2017; revised on March 28, 2018; editorial decision on April 30, 2018; accepted on May 1, 2018

Abstract

Summary: The backbone of most proteins forms an open curve. To study their entanglement, a common strategy consists in searching for the presence of knots in their backbones using topological invariants. However, this approach requires to close the curve into a loop, which alters the geometry of curve. Knoto-ID allows evaluating the entanglement of open curves without the need to close them, using the recent concept of knotoids which is a generalization of the classical knot theory to open curves. Knoto-ID can analyse the global topology of the full chain as well as the local topology by exhaustively studying all subchains or only determining the knotted core. Knoto-ID permits to localize topologically non-trivial protein folds that are not detected by informatics tools detecting knotted protein folds.

Availability and implementation: Knoto-ID is written in C++ and includes R (www.R-project.org) scripts to generate plots of projections maps, fingerprint matrices and disk matrices. Knoto-ID is distributed under the GNU General Public License (GPL), version 2 or any later version and is available at https://github.com/sib-swiss/Knoto-ID. A binary distribution for Mac OS X, Linux and Windows with detailed user guide and examples can be obtained from https://www.vital-it.ch/software/Knoto-ID.

Contact: julien.dorier@sib.swiss

1 Introduction

The observation that protein backbones can form knots (Mansfield, 1994) initiated numerous studies of their nature and potential advantages or disadvantages that they may provide (e.g.Dabrowski-Tumanski *et al.*, 2016; Virnau *et al.*, 2006). In this context, it was important to classify protein knots in terms of their topology.

A knot is a closed curve in 3-dimensional space that doesn't intersect itself (Adams, 1994). However, the backbones of many biomolecules and specifically of many proteins correspond to open spatial curves and so, in strict topological sense, such curves are classified as unknotted. Until recently, the only way to study the topology of an open protein chain was to first close them and then proceed with the study of the entanglement. Of course, closing the chain alters its geometry. In 2012, V. Turaev introduced the concept of knotoids as a generalization of the classical knot theory to open knots (Turaev, 2012). Knotoids were studied further by N. Gügümcü and L. H. Kauffman (Gügümcü and Kauffman, 2017). As a consequence, a number of studies emerged that implemented this new mathematical tool in the analysis of a protein backbone (Alexander, 2017; Goundaroulis *et al.*, 2017a,b). An important advantage of the knotoid approach to analyse protein structure is the possibility to detect topologically nontrivial protein folds that are not detected using the knotting approach (Goundaroulis *et al.*, 2017a).

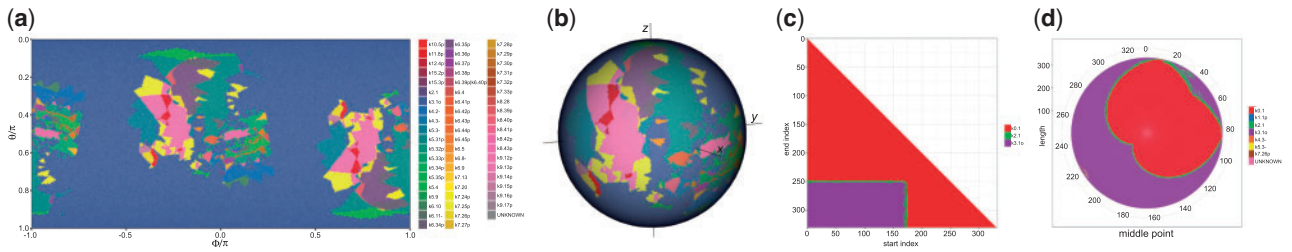

Fig. 1. (**a**) Projection map on a plane using spherical co-ordinates or (**b**) on a globe using Cartesian co-ordinates. (**c**) Fingerprint matrix and (**d**) a disc matrix. This figure was generated with Knoto-ID using the protein 3KZN (Shi *et al.*, 2006) as input

In this note, we introduce Knoto-ID, a command line tool that is able to analyse and classify open spatial curves using this new mathematical concept. Moreover, we provide the possibility of closing an open 3D curve, if a knot analysis is required, with either a direct closure or using the uniform closure technique (e.g. Sulkowska *et al.*, 2012). This note focuses on individual open protein chains, however Knoto-ID can be used to analyse any open linear conformation in 3-space such as chromosomes (Siebert *et al.*, 2017), synthetic polymers, random walks, etc.

2 Implementation

To analyse a protein, the co-ordinates of the Cα atoms of the protein backbone have to be extracted from a pdb file downloaded from the PDB (www.rcsb.org, Berman *et al.*, 2000) and then stored to a text file, which is the input of Knoto-ID. The backbone is then placed inside a large enough enclosing sphere. Each point of the sphere defines a direction of projection. For a given direction of projection, two infinite lines are introduced that pass from each of the termini of the curve and are parallel to the chosen direction. A triangle elimination method based on the KMT algorithm (Koniaris and Muthukumar, 1991; Taylor, 2000) is then applied to simplify the curve while preserving its underlying topology with respect to the two parallel lines. A knotoid diagram is obtained by projecting the curve on an oriented surface (a plane or a sphere). Finally, a topological invariant is evaluated on the knotoid diagram. For curves projected on a sphere the topological invariant is the Jones polynomial for knotoids (Turaev, 2012), while for curves projected on a plane it is the Turaev loop bracket polynomial (Turaev, 2012; see the Knoto-ID user guide for a brief description of the theory). The knotoid type corresponding to the resulting polynomial can be optionally obtained using a list knotoid types distributed with Knoto-ID. Different choices of projection directions may yield different diagrams and so one has to sample an adequate number of projection directions in order to approximate the spectrum of knotoid types that corresponds to the spatial curve. The spectrum of knotoid types can be visualized using projection maps generated by Knoto-ID (Fig. 1a and b). Moreover, Knoto-ID is also able to handle closed chains as input, or to create a closed chain from an open one using either direct or uniform closure. Here, the classical Jones polynomial is used (Jones, 1987).

Finally, Knoto-ID can analyse all subchains of a given curve to produce a fingerprint matrix (Jamroz *et al.*, 2015; King *et al.*, 2007), for the case of open chains or a disc matrix (Rawdon *et al.*, 2015), for the case of closed chain (Fig. 1c and d). In addition, Knoto-ID can find the knotted core of the chain, which is the shortest subchain obtained by progressively trimming the input chain without changing the dominant knot or knotoid type in the process.

3 Conclusion

Knoto-ID is the first tool that is able to handle, analyse as well as classify open linear conformations in 3-space such as proteins in terms of their topology without requiring them to be closed into a loop, using the concept of knotoids.

Acknowledgements

The authors thank Louis H. Kauffman for fruitful discussions, Frédéric Schütz for his advice on Knoto-ID packaging. They also thank Eric Rawdon, Elizabeth Annoni and Nicole Lopez for kindly providing the list of projections distributed with Knoto-ID.

Funding

The work was funded in part by Leverhulme Trust (RP2013-K-017) and by the Swiss National Science Foundation (31003A_166684), both credited to Andrzej Stasiak.

Conflict of Interest: none declared.

References

Adams,C.C. (1994) *The Knot Book*. Freeman and Company, New York.
Alexander,K. (2017) Proteins analysed as virtual knots. *Sci. Rep.*, **7**, 42300.
Berman,H.M. *et al.* (2000) The protein data bank. *Nucleic Acids Res.*, **28**, 235–242.
Dabrowski-Tumanski,P. *et al.* (2016) In search of functional advantages of knots in proteins. *PLoS One*, **11**, e0165986.
Goundaroulis,D. *et al.* (2017a) Studies of global and local entanglements of individual protein chains using the concept of knotoids. *Sci. Rep.*, **7**, 6309.
Goundaroulis,D. *et al.* (2017b) Topological models for open-knotted protein chains using the concepts of knotoids and bonded knotoids. *Polymers*, **9**, 444.
Gügümcü,N. and Kauffman,L.H. (2017) New invariants of knotoids. *Eur. J. Combin.*, **65**, 186–229.
Jamroz,M. *et al.* (2015) KnotProt: a database of proteins with knots and slipknots. *Nucleic Acids Res.*, **43**, D306–D314.
Jones,V.F.R. (1987) Hecke algebra representations of braid groups and link polynomials. *Ann. Math.*, **126**, 335–388.
King,N.P. *et al.* (2007) Identification of rare slipknots in proteins and their implications for stability and folding. *J. Mol. Biol.*, **373**, 153–166.
Koniaris,K. and Muthukumar,M. (1991) Self-entanglement in ring polymers. *J. Chem. Phys.*, **95**, 2873–2881.
Mansfield,M.L. (1994) Are there knots in proteins? *Nat. Struct. Biol.*, **1**, 213–214.
Rawdon,E.J. *et al.* (2015) Subknots in ideal knots, random knots, and knotted proteins. *Sci. Rep.*, **5**, 8298.

Shi,D. *et al*. (2006) Structures of N-acetylornithine transcarbamoylase from Xanthomonas campestris complexed with substrates and substrate analogs imply mechanisms for substrate binding and catalysis. *Proteins: Struct. Funct. Bioinf.*, **64**, 532–0134.

Siebert,J. *et al*. (2017) Are there knots in chromosomes? *Polymers*, **9**, 317.

Sulkowska,J.I. *et al*. (2012) Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Natl. Acad. Sci. USA*, **109**, E1715.

Taylor,W.R. (2000) A deeply knotted protein structure and how it might fold. *Nature*, **406**, 916.

Turaev,V. (2012) *Knotoids. Osaka J. Math.*, **49**, 195–223.

Virnau,P. *et al*. (2006) Intricate knots in proteins: function and evolution. *PLOS Comput. Biol.*, **2**, 1074–1079.