BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 10 2012, pages 1400–1401
doi:10.1093/bioinformatics/bts160

Structural bioinformatics

Advance Access publication April 5, 2012

Rknots: topological analysis of knotted biopolymers with R

Federico Comoglio[1,2,*] and Maurizio Rinaldi[3]

[1]Department of Biosystems Science and Engineering, ETH Zurich, Basel, CH-4058, Switzerland, [2]Life Science Zurich PhD Program on Systems Biology and [3]Dipartimento di Scienze del Farmaco di Novara, University of Piemonte Orientale "A. Avogadro", Novara, 28100, Italy

Associate Editor: Anna Tramontano

ABSTRACT

Motivation: Rknots is a flexible R package providing tools for the detection and characterization of topological knots in biological polymers. The package is well documented and provides a simple syntax for data import and preprocessing, structure reduction, topological analyses and 2D and 3D visualization. Remarkably, Rknots is not limited to protein knots and allows researchers from interdisciplinary fields to analyze different topological structures and to develop simple yet fully custom pipelines.

Availability: Rknots is distributed under the GPL-2 license and is available from the CRAN (the Comprehensive R Archive network) at http://cran.r-project.org/web/packages/Rknots

Contact: federico.comoglio@bsse.ethz.ch

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on December 24, 2011; revised on February 27, 2012; accepted on March 31, 2012

1 INTRODUCTION

The topological study of knotted biological polymers is an active interdisciplinary field of research (Bölinger *et al*., 2010; Marenduzzo *et al*., 2009; Shakhnovich, 2011). In polymers, knots influence both material properties and polymer chain dynamics (Koniaris and Muthukumar, 1991). Informally, knots are closed curves in 3D and links are collections of non-intersecting knots. Hence, knot theory deals with closed structures (see Supplementary Material for an informal introduction to the subject). However, the vast majority of biological or synthetic polymers are open chains. In this context, the definition of knot is relaxed and transferred to open curves. A chain is knotted if it does not disentangle after being pulled from both ends.

Nature tends to avoid knots. Knotted protein backbones are rare (Taylor, 2007; Virnau *et al*., 2006) and the physical mechanisms governing their formation is largely unknown. Intriguing experimental results unraveling protein knots properties have recently been reported (Sayre *et al*., 2011; Virnau *et al*., 2011) and in certain viral capsids there are evidences of mechanisms preventing knotted DNA formation (Burnier *et al*., 2008). In order to understand structural properties of knotted polymers, rigorous yet simply to use, generalized computational methods are required. Web servers for protein knots analysis (Kolesov *et al*., 2007; Lai *et al*., 2007) nicely accomplished this for proteins. However, the underlying framework cannot be generalized to cope with more complex structures and it cannot be customized.

In this note we present Rknots, an R package combining a generalized framework for the topological analysis of knotted biopolymers with the benefits of R programming. Different structures can be analyzed with a simple syntax and methods have been implemented accounting for modularity. Rknots requires a standard R installation and depends on bio3d (Grant *et al*., 2006 available from http://mccammon.ucsd.edu/~bgrant/bio3d/), rgl [Adler,D. and Murdoch,D. (2011) rgl: 3D visualization device system (OpenGL)] and rsympy [Grothendieck,G. *et al*. (2010) rSymPy: R interface to SymPy computer algebra system. R package version 0.2-1.1]. In the following we will provide an overview of the available methods. A case study on a knotted protein will be used as an example. Additional examples can be found in the package manual and vignette at http://cran.r-project.org/web/packages/Rknots

2 APPROACH

Proteins can be loaded in `.pdb` format from the file system or by fetching the Protein Data Bank (PDB; Berman *et al*., 2000) and they undergo a dedicated preprocessing (see Supplementary Material). Coordinates are then stored in the S4 `Knot` class (see the package vignette for details) and the following workflow applies afterwards. First, open chains are closed and a knot diagram is obtained through the here proposed principal component analysis projection (PCAP) algorithm (see Supplementary Material). Figure 1A illustrates a protein diagram obtained with PCAP, whereas Figure 1B–D illustrate the results of a simulation on 1000 proteins sampled randomly from the PDB in comparison to a set of generic projections or to standard projection (see Supplementary Material for details). Second, the structure is reduced to the minimal set of points topologically equivalent to the original structure, by applying a reduction algorithm. This step removes structural redundancies, speeding up downstream computations. Two structure reduction algorithms, Alexander-Briggs (Alexander and Briggs, 1926) and MSR (Comoglio and Rinaldi, 2011) are implemented in Rknots. The former has been proposed for the first time in polymers knot theory by Koniaris and Muthukumar (Comoglio and Rinaldi, 2011; Koniaris and Muthukumar, 1991). We also provide an extension of this algorithm for links.

Finally, knots can be detected in the minimal structure by computing a topological invariant, generally a polynomial. Protein knots discovered so far are among the simplest types (Bölinger *et al*., 2010; Virnau *et al*., 2006). The Alexander polynomial is sufficient to

*To whom correspondence should be addressed.

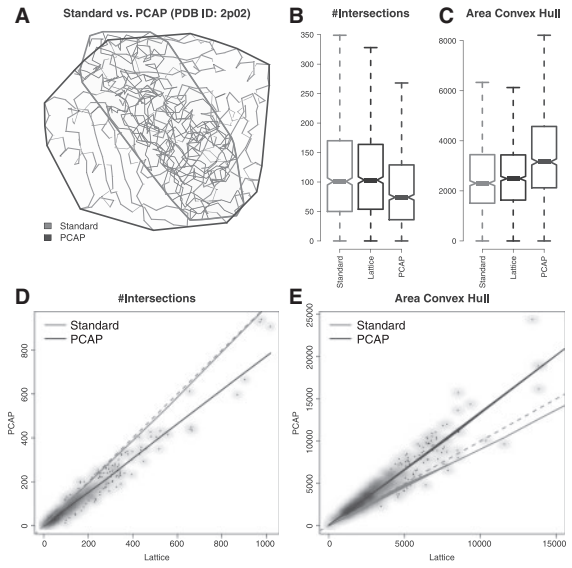

Fig. 1. The PCAP algorithm significantly improves the diagram resolution. (**A**) Diagrams obtained through either Standard projection or PCAP for the PDB ID 2p02. (**B**, **E**) The area of the diagram convex hull is significantly increased by PCAP with respect to Standard projection or to a generic set of projections (Lattice). (**C**, **D**) The number of crossings of the diagram is significantly reduced with respect to Standard projection or Lattice. See Supplementary Material for details.

characterize them unambiguously. Being its computation faster than other polynomial invariants, the Alexander polynomial is preferred when a large number of structures has to be tested. A normalized multivariable version of the Alexander polynomial (MVA) is also implemented in Rknots (see Supplementary Material) and it can be applied to the analysis of link structures such as catenanes. Rknots is not limited to the analysis of proteins. The Jones and the HOMFLY polynomial have been implemented to analyze more complex structures. Additionally, they are able to determine knots chirality. The next section shows how to apply the described workflow to the analysis of a protein knot.

3 APPLICATION

The following example illustrates a typical Rknots session for the analysis of a protein of interest. A monomeric left-handed trefoil knotted protein (PDB ID 2k0a) has been selected for this case study. First, the protein is imported with the function `loadProtein`. Gaps finding is performed with default parameter (`cutoff = 7`):

```
library(Rknots)
pdb <- loadProtein(pdbID = "2K0A", cutoff = 7)
```

Second, a `Knot` object is created simply by providing the protein coordinates of the single polypeptide chain to `newKnot`:

```
chain <- newKnot(pdb$A)
```

Third, the open chain is closed and projected using the `closeAndProject` function. The knot diagram can then be visualized with `plot`:

```
chain <- closeAndProject(chain)
plot(chain, lwd = 1.5)
```

Then, the knot type is determined by computing a polynomial invariant with the `computeInvariant` function. By setting `invariant = "HOMFLY"`, the HOMFLY polynomial is returned:

```
computeInvariant(chain, invariant = "HOMFLY")
```

Finally, further information on the knot type can be obtained with the function `getKnotType` (see the package vignette for details).

4 CONCLUSION

Rknots is the first package providing generalized tools for the study of knotted biopolymers. The modularity of the package allows integration in custom pipelines. We encourage contributions from other members of the research community.

Funding: This work was partially funded by the Italian Ministry of Education, University and Research (MIUR).

Conflict of Interest: none declared.

REFERENCES

Alexander,J.W. and Briggs,G.B. (1926) On types of knotted curves. *Ann. Math.*, **28**, 562–586.

Berman,H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, 28, 235–242.

Bölinger,D. *et al*. (2010) A Stevedore's protein knot. *PLoS Comput. Biol.* **6**, e1000731.

Burnier,Y. *et al*. (2008) DNA supercoiling inhibits DNA knotting, *Nucleic Acids Res.*, **36**, 4956–4963.

Comoglio,F. and Rinaldi,M. (2011) A topological framework for the computation of the HOMFLY polynomial and its application to proteins. *PLoS ONE*, **6**, e18693.

Grant,B. *et al*. (2006) Bio3D: an R package for the comparative analysis of protein structures. *Bioinformatics*, **22**, 2695–2696.

Kolesov,G. *et al*. (2007) Protein knot server: detection of knots in protein structures. *Nucleic Acid Res.*, **35**, W425–W428.

Koniaris,K. and Muthukumar,M. (1991) Knottedness in ring polymers. *Phys. Rev. Lett.*, **66**, 2211–2214.

Lai,Y.L. *et al*. (2007) pKNOT: the protein KNOT web server. *Nucleic Acids Res.*, **35**, W420–4

Marenduzzo,D. *et al*. (2009) DNA-DNA interactions in bacteriophage capsids are responsible for the observed DNA knotting. *Proc. Natl. Acad. Sci. USA*, **106**, 22269–22274

Sayre,T.C. *et al*. (2011) Protein stabilization in a highly knotted protein polymer. *Protein Eng. Des. Sel.*, **24**, 627–630.

Shakhnovich,E. (2011) Protein folding: to knot or not to knot? *Nat. Mater*, **10**, 84–86.

Taylor,W.R. (2007) Protein knots and fold complexity: some new twists. *Comput. Biol. Chem.*, **31**, 151–162.

Virnau,P. *et al*. (2006) Intricate knots in proteins: function and evolution. *PLoS Comput. Biol.*, **2**, e122.

Virnau,P. *et al*. (2011) Structures and folding pathways of topologically knotted proteins. *J. Phys. Conds. Matter*, **23**, 033101.