Protein knot server: detection of knots in protein structures

Citation

Published Version
doi:10.1093/nar/gkm312

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:4874789

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Protein knot server: detection of knots in protein structures
Grigory Kolesov1, Peter Virnau2,3,*, Mehran Kardar2 and Leonid A. Mirny1,2,*

1Harvard–MIT Division of Health Sciences and Technology, 2Department of Physics, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA and 3Institut für Physik, Johannes Gutenberg-Universität Mainz, Staudinger Weg 7, 55099 Mainz, Germany

Received January 31, 2007; Revised April 1, 2007; Accepted April 14, 2007

ABSTRACT

KNOTS (http://knots.mit.edu) is a web server that detects knots in protein structures. Several protein structures have been reported to contain intricate knots. The physiological role of knots and their effect on folding and evolution is an area of active research. The user submits a PDB id or uploads a 3D protein structure in PDB or mmCIF format. The current implementation of the server uses the Alexander polynomial to detect knots. The results of the analysis that are presented to the user are the location of the knot in the structure, the type of the knot and an interactive visualization of the knot. The results can also be downloaded and viewed offline. The server also maintains a regularly updated list of known knots in protein structures.

INTRODUCTION

Interest in the topological properties of biological systems was greatly accelerated with the discovery of knots in single-stranded DNA in 1976 (1). Subsequently, knots in DNA were investigated extensively (2–5) and even created artificially in polymeric materials (6), but it took another 20 years before the first systematic studies of protein knots appeared (7–11). Topology is particularly relevant for proteins because the 3D structure of a protein directly determines its functionality. Recently, we performed a comprehensive analysis of the Protein Data Bank (11) and demonstrated that knotted structures tend to persist across species and kingdoms. However, when a knot appears or vanishes in the course of evolution, the function of the protein is also altered accordingly (11–13). We uncovered some knotted proteins that have significant biomedical importance, such as the Parkinson’s disease-associated ubiquitin hydrolase UCH-L1 (14) or its structural homolog UCH-L3 (10,15), which contain the most complicated knots found in proteins so far. Other challenges include understanding the folding and unfolding of knotted proteins. The underlying mechanisms are not yet well understood and are the subject of active research (16,17).

Surprisingly, most discovered knots were not reported at the time the structure was solved, since finding knots in protein structures by naked eye is virtually impossible. Moreover, widely used protein structure verification tools like WHATIF (18), VERIFY3D (19) and PROCHECK (20) do not have the capability to detect knots. We hope that with our contribution, the discovery of knots in newly solved protein structures becomes part of the standard routine, similar to identification of secondary structure elements or classification of protein’s architecture.

To address this challenge, we developed a web server that allows a user to check a new or a known protein structure for knots by entering its PDB id or uploading a coordinate file.

MATERIALS AND METHODS

How knots are determined

Mathematically, knots are only well defined in closed (circular) loops (21). However, both the N- and C-termini of open proteins are typically located close to the surface of the protein and can be connected unambiguously: We reduce the protein to its backbone and draw two lines outward starting at the termini in the direction of the connection line between the center of mass of the backbone and the respective ends. The two lines are joined by a big loop, and the structure is topologically classified by the computation of its Alexander polynomial (21,22). To determine an estimate for the size of the knotted core, we successively delete amino acids from the N-terminus until the protein becomes unknotted (11). The procedure is repeated at the C-terminus starting with the last N-terminal deletion structure that contained the original knot. For each deletion, the outward-pointing line through the new termini is parallel to the respective lines...
computed for the full structure. Unfortunately, the size of a knot is not always precisely determined by this procedure, so reported sizes should only be regarded as approximate.

To speed up calculations, the KMT reduction scheme is used (9,11,23,24). This algorithm successively deletes amino acids that are not essential to the topological structure of the protein. It is also employed to create a reduced representation of the knot (Figure 1).

In the course of our investigations (11) we came up with a number of stringent criteria that a structure should satisfy to be classified as knotted:

(i) The Alexander polynomial should yield a knot.
(ii) There should not be any gaps in the polypeptide backbone. (See below.)
(iii) The knot should persist if two amino acids are removed from each end. (This prevents knots formed by just a few residues at the end of the chain passing through the loop—‘shallow knots’ and knots which only appear due to our specific loop-closure procedure.)

Unfortunately, there are some structures containing regions of the backbone that were not resolved and for which coordinates are not reported in PDB (a gap in the structure). Mobile loops may not be resolved by X-ray crystallography unless they are stabilized by a ligand or by protein engineering, for example. If the polypeptide chain contains a gap, the knot is reported if (i) a knot is present in at least one fragment of the chain and (ii) the structure that results from gaps being bridged with straight lines contains a knot. These criteria form the basis of our list of known knots. We have also included knotted structures with gaps if at least one homolog is knotted.

A Knots found in the 1uam structure:

<table>
<thead>
<tr>
<th>Knot residues</th>
<th>Chain start-stop</th>
<th>Knot type</th>
<th>Knot</th>
</tr>
</thead>
</table>
| 86-130A       | -1-250A          | 31 (trefoil knot) | ![Image](image)

Download results and rasmol scripts as zip package

B Residues 86-130A

<table>
<thead>
<tr>
<th>Knot in the 1uam structure</th>
<th>Simplified representation of the knot</th>
</tr>
</thead>
</table>

![Image](image)

Hide/Show unknotted structure Spin the structures

Hint: hold Ctrl-Alt to move the structure, Shift to zoom. Right click to get console.
Knotted region is defined as ‘knot’, typing ‘select knot’ will select corresponding residues.

Enter one-line RasMol/Chime script commands here:

Figure 1. The output of the Knots server for *H. influenzae* TrmD (PDB id 1uam). (A) Page one: the summary table. (B) Page two: Jmol interactive visualization. The 1uam structure is displayed in the left window with a knot highlighted in rainbow colors and the rest of the protein hidden. In this case, the trefoil knot spans a relatively small region of the protein and can be easily seen by eye in the protein structure. In many cases, this is difficult and the right panel offers the view of a simplified (reduced) representation of the knot. These visualizations can also be viewed offline using Rasmol scripts provided in the downloadable package.
that unlike in our previous work (11), PDB residue numbers are used to describe the location of the knots. From each terminus; in parentheses we indicate how many amino acids can be removed from each side before the structure becomes unknotted. Note

Ubiquitin hydrolase UCH-L1
Homo sapiens

Table 1. Protein knots discovered in 2006

<table>
<thead>
<tr>
<th>Protein</th>
<th>Species</th>
<th>PDB code</th>
<th>Length</th>
<th>Type</th>
<th>Knotted core</th>
</tr>
</thead>
<tbody>
<tr>
<td>α/β knot</td>
<td>Homo sapiens</td>
<td>2ha8</td>
<td>159</td>
<td>3₁</td>
<td>103–148 (30)</td>
</tr>
<tr>
<td>S-Adenosylmethionine synthetase</td>
<td>Porphyromonas gingivalis</td>
<td>2i6d</td>
<td>231</td>
<td>3₁</td>
<td>177–222 (9)</td>
</tr>
<tr>
<td>Ubiquitin hydrolase UCH-L1</td>
<td>Homo sapiens</td>
<td>2p02</td>
<td>380</td>
<td>3₁</td>
<td>59–302 (21)</td>
</tr>
</tbody>
</table>

Length refers to the size of the protein in amino acids. The knotted core is the minimum configuration that stays knotted after a series of deletions from each terminus; in parentheses we indicate how many amino acids can be removed from each side before the structure becomes unknotted. Note that unlike in our previous work (11), PDB residue numbers are used to describe the location of the knots.
CONCLUSION AND OUTLOOK

In this article, we presented our knot detection server and an illustration of its use. The server is easy to use, accurate and fast. In future, we plan to add automatic modeling of unresolved parts in the structures by using homology.

ACKNOWLEDGEMENTS

This work was supported by National Science Foundation grant DMR-04-26677 and by the Deutsche Forschungsgemeinschaft grant VI237/1. L.M. is an Alfred P. Sloan Research Fellow. Funding to pay the Open Access publication charges for this article was provided by the NIH-funded National Center for Biomedical Computing, Informatics for Integrating Biology and the Bedside (i2b2).

Conflict of interest statement. None declared.

REFERENCES