Protein knot server: detection of knots in protein structures

Protein knot server: detection of knots in protein structures

Grigory Kolesov[1], Peter Virnau[2,3,*], Mehran Kardar[2] and Leonid A. Mirny[1,2,*]

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

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:

Knot residues	Chain start-stop	Knot type	Knot	
86-130A	-1-250A	31 (trefoil knot)		Jmol visualization

Download results and rasmol scripts as zip package

B Residues 86-130A

Knot in the 1uam structure	Simplified representation of the knot
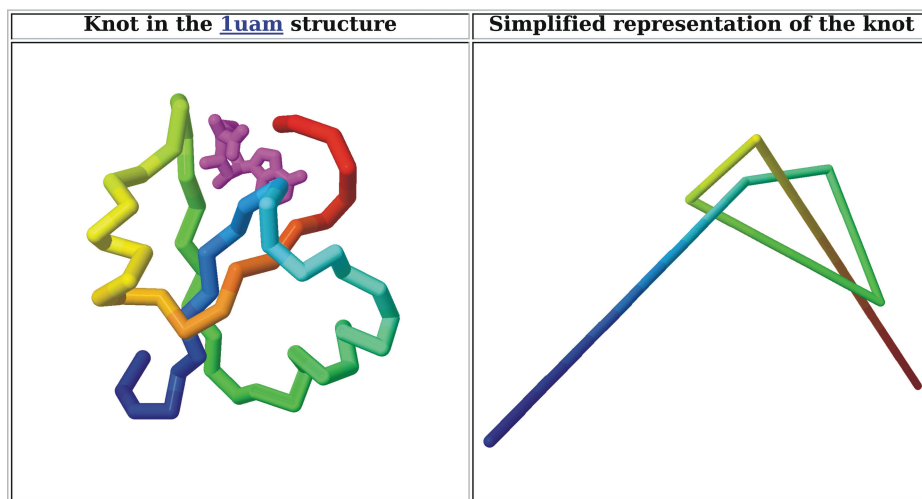
Hide/Show unknotted structure | Spin the structures

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.

Server: input and output

As an input, the server accepts a structure file (in PDB or mmCIF format) or a PDB ID. The structure is tested for knots as described above. An option allowing a user to decide how to deal with the unresolved part of the structure is provided. The user may choose to connect unresolved parts by straight lines or to treat them as described above.

Figure 1 presents a typical output of the server—the summary page reporting a knot. If a knot is found, the server reports the type of the knot (e.g. 3_1- the trefoil knot, 4_1- the 'figure eight' knot, 5_2, etc.), its location in the protein structure, and a simplified representation of the knot (Figure1A). At this point, a user may choose to download results of the calculation as a collection of Rasmol/Jmol scripts or to proceed to the second page that has Jmol visualization of the knot on our server.

An option to download the tabulated results, the original structure file, and simplified structure and visualization Rasmol/Jmol scripts in one zip package is provided. This also can be used in cases when Jmol fails to start due to structure size or web browser-Java incompatibility issues.

The second page (Figure1B) has a two-window GUI to examine, rotate and further analyze the structure of the knot. The left window visualizes the protein structure with the knot using a Jmol Java applet. The knotted part is colored in rainbow colors to facilitate following the chain and visualizing the knot. The right window presents a simplified representation of the knot obtained by the reduction algorithm, making it easier to see that the protein structure is indeed knotted. The structures in both windows can be rotated, magnified and further analyzed using the tools of Jmol applet. Two buttons below (i) hide or show the rest of the protein structure in the left window, thus allowing a user to focus on the knot or to examine it in the context of the structure; and (ii) to spin structures in both windows simultaneously. An expert user familiar with Rasmol/Jmol commands can further analyze the structure using the command-line interface by entering individual commands or a whole script into a field below the windows.

The front page of the server also provides a curated list of discovered knots in proteins, classified according to the type of the knot, as well as a brief definition.

The server is implemented as a CGI Perl script, while the algorithmic part is written in C. The results of the calculation are stored for 20 min on the server, after which they have to be recomputed. Knot detection typically takes one to a few seconds.

Example of using the server

The bacterial tRNA(m^1G37)methyltransferase (TrmD) is an enzyme that transfers methyl group from s-adenosyl-L-methionine (AdoMet) to a G nucleotide in the anti-codon region of certain bacterial tRNA species. The methylation of anti-codon nucleotides is essential for reducing the error rate in anti-codon binding to the complementary codon on mRNA during translation. The crystal structure of the enzyme from *Haemophilus influenzae* has recently been solved and is known to have a trefoil knot in the AdoMet-binding pocket (26). The specific configuration of the pocket allows AdoMet to adopt an unusual strongly bent conformation with its methyl group protruding from the pocket and accessible for transfer reaction (26).

The Knots output for a PDB entry 1uam, *H. influenzae* TrmD protein, is shown in Figure 1. A trefoil knot has been correctly identified for residues 86–130A in the 1uam structure. Clicking on 'Jmol visualization' link leads to the second page showing a protein ribbon diagram (Figure 1B, left), and the simplified representation of the knot. The knot can be easily seen in the protein structure by eye if the surrounding structure is hidden from view using the button provided.

The reduced representation of the knot (Figure 1B, right panel) is generated by the KMT reduction algorithm. The first and the last segments in this representation are not part of the protein but represent the connection lines to 'infinity', which are required to circularize the structure and calculate the Alexander polynomial (see the Section 'How knots are determined').

To ensure that the knot is not an artifact of connecting a gap in the structure, one may want to test for knots in each protein fragment separately. This option is provided on the front page of the server. In the case of the 1uam structure, the knot is found in one of the fragments.

More examples of knot in protein structure and their analysis can be found in our recent publication (11).

New protein knots discovered in 2006

Table 1 lists all novel protein knots that were discovered with our software in 2006. A complete list with all knotted proteins is available online.

Table 1. Protein knots discovered in 2006

Protein	Species	PDB code	Length	Type	Knotted core
α/β knot	*Homo sapiens*	2ha8	159	3_1	103–148 (30)
	Porphyromonas gingivalis	2i6d	231	3_1	177–222 (9)
s-Adenosylmethionine synthetase	*Homo sapiens*	2p02	380	3_1	59–302(21)
Ubiquitin hydrolase UCH-L1	*Homo sapiens*	2etl	219	5_2	10–216 (7)

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.

REFERENCES

1. Liu,L.F., Depew,R.E. and Wang,J.C. (1976) Knotted single-stranded DNA rings: a novel topological isomer of circular single-stranded DNA formed by treatment with Escherichia coli omega protein. *J. Mol. Biol.*, **106**, 439–452.
2. Dean,F.B., Stasiak,A., Koller,T. and Cozzarelli,N.R. (1985) Duplex DNA knots produced by *Escherichia coli* topoisomerase I. *J. Biol. Chem.*, **260**, 4975–4983.
3. Rybenkov,V.V., Cozzarelli,N.R. and Vologodskii,A.V. (1993) The probability of DNA knotting and the effective diameter of the DNA double helix. *Proc. Natl Acad. Sci. USA*, **90**, 5307–5311.
4. Shaw,S.Y. and Wang,J.C. (1993) Knotting of a DNA chain during ring closure. *Science*, **260**, 533–536.
5. Arsuaga,J., Vazquez,M., Trigueros,S., Sumners,D.W. and Roca,J. (2002) Knotting probability of DNA molecules confined in restricted volumes: DNA knotting in phage capsids. *Proc. Natl Acad. Sci. USA*, **99**, 5373–5377.
6. Lukin,O. and Vögtle,F. (2005) Knotting and threading of molecules: chemistry and chirality of molecular knots and their assemblies. *Angew. Chem. Int. Ed.*, **44**, 1456–1477.
7. Mansfield,M.L. (1994) Are there knots in proteins? *Nat. Struct. Mol. Biol.*, **1**, 213–214.
8. Mansfield,M.L. (1997) Fit to be tied. *Nat. Struct. Mol. Biol.*, **4**, 166–167.
9. Taylor,W.R. (2000) A deeply knotted protein structure and how it might fold. *Nature*, **406**, 916–919.
10. Lua,R.C. and Grosberg,A.Y. (2006) Statistics of knots, geometry of conformations, and evolution of proteins. *PLoS Comp. Biol.*, **2**, 350–357.
11. Virnau,P., Mirny,L.A. and Kardar,M. (2006) Intricate knots in proteins: function and evolution. *PLoS Comp. Biol.*, **2**, 1074–1079.
12. Morizono,H, Cabrera-Luque,J., Shi,D., Gallegos,R., Yamaguchi,S., YuX.L.,, Allewell,N.M.,, Malamy,M.H., and Tuchman,M., (2006) Acetylornithine transcarbamylase: a novel enzyme in arginine biosynthesis. *J. Bacteriol.*, **188**, 2974–2982.
13. Shi,D., Morizono,H., Aoyagi,M., Tuchman,M. and Allewell,N.M. (2000) Crystal structure of human ornithine transcarbamylase complexed with carbamyl phosphate and L-norvaline at 1.9 A resolution. *Proteins: Struct. Funct. Genet.*, **39**, 271–277.
14. Das,C., Hoang,Q.Q., Kreinbring,C.A., Luchansky,S.J., Meray,R.K., Ray,S.S., Lansbury,P.T., Ringe,D. and Petsko,G.A. (2006) Structural basis for conformational plasticity of the Parkinson's disease-associated ubiquitin hydrolase UCH-L1,. *Proc. Natl Acad. Sci. USA*, **103**, 4675–4680.
15. Misaghi,S., Galardy,P.J., Meester,W.J.N., Ovaa,H., Ploegh,H.L. and Gaudet,R., (2005) Structure of the ubiquitin hydrolase Uch-L3 complexed with a suicide substrate. *J. Biol. Chem.*, **280**, 1512–1520.
16. Mallam,A.L. and Jackson,S.E. (2005) Folding studies on a knotted protein. *J. Mol. Biol.*, **346**, 1409–1421.
17. Mallam,A.L. and Jackson,S.E. (2006) Probing nature's knots: the folding pathway of a knotted homodimeric protein,. *J. Mol. Biol.*, **359**, 1420–1436.
18. Vriend,G. (1990) WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.*, **8**, 52–56.
19. Eisenberg,D., Luthy,R. and Bowie,J.U. (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, **277**, 396–404.
20. Laskowski,R.A., MacArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) . PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283–291.
21. Adams,C.C. (1994) *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. Freeman, New York.
22. 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.
23. Virnau,P., Kardar,M. and Kantor,Y. (2005) Capturing knots in polymers. *Chaos*, **15**, 041103.
24. Koniaris,K. and Muthukumar,M. (1991) Self-entanglement in ring polymers. *J. Chem. Phys.*, **95**, 2873–2881.