Knot theory and proteins

Isabel K. Darcy
University of Iowa
www.math.uiowa.edu/~idarcy

Intricate Knots in Proteins: Function and Evolution
Rapid knot detection and application to protein structure prediction.
Firas Khatib, Matthew T. Weirauch and Carol A. Rohl, Bioinformatics. 2006 Jul 15;22(14)

The Knotfind algorithm is available in the Rosetta structure prediction program at http://www.rosettacommons.org and the Undertaker program (Karplus et al 2005)
Ordering triples:

1.) use the distance between vertex i-1 and i+1. Start with largest distance.

2.) use area of triangle formed by vertex i-1, i, i+1. Start with largest area.
State is trapped if triples ordered by distance.

Only false positive in a check of 9,557 proteins from RCSB PDB.

(< 90% sequence identity, x-ray res < 3Å, no R-factor filtering + 4)

Conformation becomes unknotted if triples ordered by area

Area: 4 false positives
N to C: 7 false positives.
21 deeply knotted proteins + 18 shallow proteins found.

Solid lines = sequence similarity
(BLASTp evalue < 1E – 05)

Dotted lines = Structural similarity
(MAMMOTH evalue < 1E -07)
A deeply knotted protein structure and how it might fold
pKNOT: the protein KNOT web server
50 iterations usually sufficient. Sometimes 500+ iterations needed.

Knotted proteins come from the following classes:
(1) methyltransferase,
(2) transcarbamylase,
(3) carbonic anhydrase,
(3) ketol–acid reductoisomerase,
(4) ubiquitin hydrolase,
(5) methionine adenosyl transferase,
(6) the chromophore-binding domain of bacterial phytochrome and
(7) the inner core shell component protein of bluetongue virus.

In addition, we also identified two knotted NMR structures: 1POQ and 1J2O. However, it is not clear whether these knots are authentic or due to incorrect structural refinement, since only one knotted model is identified among all NMR models for each protein (model 7 in 1POQ and model 14 in 1J2O).
The proteins with a trefoil knot are
(1) methyltransferase,
(2) transcarbamylase,
(3) methionine adenosyltransferase,
(4) carbonic anhydrase and
(5) YMPa superantigen (NMR).

The proteins with a 4.1 knot are
(1) the chromophore-binding domain of bacterial phytochrome,
(2) the core protein of bluetongue virus,
(3) ketol–acid reductoisomerase and
(4) a LIM-Ldbl-LID chimeric protein (NMR).

The only protein family with a 5.2 knot is ubiquitin c-terminal hydrolase (1).
Linear Random Knots and Their Scaling Behavior
Kenneth Millett, Akos Dobay, and Andrzej Stasiak,
Linear Random Knots and Their Scaling Behavior
Kenneth Millett, Akos Dobay, and Andrzej Stasiak,

10 000 random closures
The diameter of the enclosing sphere is not to scale.
Statistics of knots, geometry of conformations, and evolution of proteins.
Rhonald C. Lua, Alexander Y. Grosberg
PLoS Comput Biol. 2006 May;2(5)

<table>
<thead>
<tr>
<th></th>
<th>unknot</th>
<th>3.1</th>
<th>4.1</th>
<th>5.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>4516</td>
<td>164</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>Center</td>
<td>4692</td>
<td>20</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Random</td>
<td>4697</td>
<td>15</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

and more complicated knots for random closure
Fraction of Protein Chains at a Given Length with a Trivial Knot (01) in the RANDOM Method, Plotted against the Length or Number of Residues. Adjacent points are connected by dashed lines.

The data for the trivial knotting probability of compact lattice loops (from $4 \times 4 \times 4$ to $12 \times 12 \times 12$) is included, shown connected by thick lines.
**Protein knot server: detection of knots in protein structures**

### 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>
<tbody>
<tr>
<td>86-130A</td>
<td>-1-250A</td>
<td><img src="image" alt="31 knot" /></td>
<td>Imol visualization</td>
</tr>
</tbody>
</table>

Download results and rasmol scripts as zip package

### B Residues 86-130A

**Knot in the 1uam structure**

**Simplified representation of the knot**

Hint: hold Ctrl+Alt to move the structure, Shift to zoom; Right click to get console. Typing 'select knot' will select corresponding residues.
Enter one-line RasMol/Chime script commands here:
Please enter pdb id (e.g. 1v2x):

Or upload file (pdb or mmCIF):

- Connect gaps in the structure by straight line
- Treat each unconnected fragment separately

Submit

**List of known knots**

<table>
<thead>
<tr>
<th>Protein</th>
<th>Species</th>
<th>PDB code</th>
<th>Length</th>
<th>Knot</th>
<th>Knotted core</th>
</tr>
</thead>
<tbody>
<tr>
<td>YbeA-like</td>
<td>E.coli</td>
<td>1ns5</td>
<td>153</td>
<td>3_1</td>
<td>69-121 (32)</td>
</tr>
<tr>
<td></td>
<td>T.maritima</td>
<td>1o6d</td>
<td>147</td>
<td>3_1</td>
<td>68-117 (30)</td>
</tr>
<tr>
<td></td>
<td>S.aureus</td>
<td>1vh0</td>
<td>157</td>
<td>3_1</td>
<td>73-126 (31)</td>
</tr>
<tr>
<td></td>
<td>B.subtilis</td>
<td>1to0</td>
<td>148</td>
<td>3_1</td>
<td>73-125 (32)</td>
</tr>
<tr>
<td>tRNA(m1G37)-methyltransferase TrmD</td>
<td>H.influenza</td>
<td>1uaj</td>
<td>241</td>
<td>3_1</td>
<td>85-130 (92)</td>
</tr>
<tr>
<td></td>
<td>E.coli</td>
<td>1p9p</td>
<td>235</td>
<td>3_1</td>
<td>90-130 (89)</td>
</tr>
<tr>
<td></td>
<td>S.cerevisiae</td>
<td>2v3k</td>
<td>219</td>
<td>3_1</td>
<td>175-225 (27)</td>
</tr>
<tr>
<td>SpoU-like RNA 2'-O ribose mtf.</td>
<td>T.thermophilus</td>
<td>1v2x</td>
<td>191</td>
<td>3_1</td>
<td>96-140 (51)</td>
</tr>
<tr>
<td></td>
<td>H.influenza</td>
<td>1j85</td>
<td>156</td>
<td>3_1</td>
<td>77-114 (42)</td>
</tr>
<tr>
<td></td>
<td>T.thermophilus</td>
<td>1ipa</td>
<td>258</td>
<td>3_1</td>
<td>190-234 (29)</td>
</tr>
<tr>
<td></td>
<td>E.coli</td>
<td>1sg0</td>
<td>242</td>
<td>3_1</td>
<td>173-215 (28)</td>
</tr>
</tbody>
</table>
Protein ends usually on surface.

Knot depth: remove amino acids from both ends until knot disappears.

Watch out for breaks in protein chains
Why are some proteins knotted?

Why are knots so rare in proteins?

1.) added stability?
   a.) degradation?
   b.) limit movement?
   c.) geometry for binding
   d.) enzymatic activity.

2.) Property of protein secondary structure?
   Unlikely, some algorithms construct knotted proteins (Khatib et. al. Bioinfomatics 2006.)

3.) Property of a protein folding method?

4.) Evolutionary selection?

4.) Detrimental only to non-enzymatic proteins??
Observation: protein knots have unknotting number 1 (Taylor 2000)
Very speculative method for creating knotted proteins (or for finding unknotted proteins for comparison).

1.) Circularize the protein backbone as described above.

2.) Take random projections of protein backbone after forming closure.

   A.) Change crossing nearest to N-terminal.

   B.) Change crossing nearest to C terminal.

3.) Determine if either conformation resulting from a crossing change in knotted.

4.) Determine distances between segments involved in desired crossing change.

Hard part: create new knotted protein -- modify unknotted protein by changing appropriate angle.
Compare statistics of number of “knotting number 1 geometric conformations with randomly generated confirmations to determine rarity of threading.
Figure 2 Structures of Transcarbamylase from *X. campestris* with a Trefoil Knot and from *Human* without a Knot

Intricate Knots in Proteins: Function and Evolution