Untangling Knots in Lattices and Proteins
A Computational Study
By Rhonald Lua
Adviser: Alexander Yu. Grosberg
University of Minnesota
Globular proteins have dense, crystal-like packing density. Proteins are small biomolecular “machines” responsible for carrying out many life processes.
Hemoglobin Protein Backbone
(string of $\alpha$–carbon units)

One chain

“ball of yarn”
4x4x4 Compact Lattice Loop

Possible cube dimensions: 2x2x2, 4x4x4, 6x6x6, ..., LxLxL, ...

No. of distinct conformations: \( \left( \frac{z-1}{e} \right)^N \) (Flory)

\[ z = 6 \text{ in 3D} \]
Hamiltonian Path Generation

(A. Borovinskiy, based on work by R. Ramakrishnan, J.F. Pekny, J.M. Caruthers)
14x14x14 Compact Lattice Loop
In this talk...

• knots and their relevance to physics
• “virtual” tools to study knots
• knotting probability of compact lattice loops
• statistics of subchains in compact lattice loops
• knots in proteins
Knot – a closed curve in space that does not intersect itself.

The first few knots:

- **Trivial knot (Unknot)**
  - 0-1

- 3-1 (Trefoil)

- 4-1 (Figure-8)

- 5-1 (Cinquefoil, Pentafoil, Solomon’s seal)
  - 5-2
Knots in Physics

• Lord Kelvin (1867): Atoms are knots (vortices) of some medium (ether).
• Knots appear in Quantum Field Theory and Statistical Physics.
• Knots in biomolecules. Example: The more complicated the knot in circular DNA the faster it moves in gel-electrophoresis experiments.
A Little Knot Math
Reidemeister’s Theorem:
Two knots are equivalent if and only if any diagram of one may be transformed into the other via a sequence of Reidemeister moves.
Compounded Reidemeister Moves

I+III

macro move

some mess

macro move

some mess
Knot Invariants - Mathematical signatures of a knot.

<table>
<thead>
<tr>
<th>Name</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alexander</td>
<td>$\Delta(-1)$</td>
</tr>
<tr>
<td>Vassiliev degree 2</td>
<td>$v_2$</td>
</tr>
<tr>
<td>Vassiliev degree 3</td>
<td>$v_3$</td>
</tr>
</tbody>
</table>

Examples:

- **Trivial knot 0-1**
  - $\Delta(-1)=1$
  - $v_2=0$
  - $v_3=0$

- **Trefoil knot 3-1**
  - $\Delta(-1)=3$
  - $v_2=1$
  - $v_3=1$
Alexander Polynomial, $\Delta(t)$
(first knot invariant/signature)

Alexander matrix for this trefoil:

\[
\begin{pmatrix}
1 & -t & t-1 \\
t-1 & 1 & -t \\
0 & t-1 & 1
\end{pmatrix}
\]

Alexander invariant: $\Delta(-1) = \text{det} \begin{pmatrix} 1 & 1 \\ -2 & 1 \end{pmatrix} = 3$
Recipe for Constructing Alexander Matrix, $a_{kj}$

$n \times n$ matrix where $n$ is the number of underpasses

In the following

index $k$ corresponds to $k$th underpass and

index $i$ corresponds to the generator number of the arc overpassing the $k$th underpass

For row $k$:

1) when $i=k$ or $i=k+1$ then

$$a_{kk}=-1, \ a_{kk+1}=1$$

2) when $i$ equals neither $k$ nor $k+1$:

If the crossing has sign -1:

$$a_{kk}=1, \ a_{kk+1}=-t, \ a_{ki}=t-1$$

If the crossing has sign +1:

$$a_{kk}=-t, \ a_{kk+1}=1, \ a_{ki}=t-1$$

3) All other elements are zero.
Gauss Code and Gauss Diagram

Gauss code for left-handed trefoil:
b - 1, a - 2, b - 3, a - 1, b - 2, a - 3

(Alternatively…)

Gauss Diagram for trefoil:

sign:

a – ‘above’
b – ‘below’
Vassiliev Invariants
(Diagram methods by M. Polyak and O. Viro)

Degree two ($v_2$): Look for this pattern:

Degree three ($v_3$): Look for these patterns:

\[ \frac{1}{2} \]

E.g. trefoil

\[ v_2 = 1 \]

\[ v_3 = 1 \]
Prime and Composite Knots

Composite knot, $K$

$K_1$ + $K_2$

Alexander: $\Delta(t)_K = \Delta(t)_{K_1} \Delta(t)_{K_2}$

Vassiliev: $v_2(K) = v_2(K_1) + v_2(K_2)$

$v_3(K) = v_3(K_1) + v_3(K_2)$
Method to Determine Type of Knot

1. Inflation/tightening for large knots.
2. Project 3D object into 2D diagram.
3. Preprocess and simplify diagram using Reidemeister moves.
4. Compute knot invariants.
5. Give object a knot-type based on its signatures.
A. Projection

3D conformation

projection process

2D knot projection

Projected nodes and links
B. Preprocessing

Using Reidemeister moves

<table>
<thead>
<tr>
<th>L</th>
<th>Average crossings before…</th>
<th>…and after reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>114</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>383</td>
<td>40</td>
</tr>
<tr>
<td>10</td>
<td>969</td>
<td>187</td>
</tr>
<tr>
<td>12</td>
<td>2057</td>
<td>537</td>
</tr>
<tr>
<td>14</td>
<td>3896</td>
<td>1219</td>
</tr>
</tbody>
</table>
C. Knot Signature Computation

| Knot   | $|\Delta(-1)|$ | $v_2$ | $|v_3|$ | Chiral? |
|--------|-------------|------|--------|--------|
| Trivial| 1           | 0    | 0      | NO     |
| 3-1    | 3           | 1    | 1      | YES    |
| 4-1    | 5           | -1   | 0      | NO     |
| 5-1    | 5           | 3    | 5      | YES    |
| 5-2    | 7           | 2    | 3      | YES    |

$|\Delta(-1)| = 3$ \text{ (Alexander)}

$v_2(T) = 1$ \text{ (Vassiliev 2)}

$|v_3(T)| = 1$ \text{ (Vassiliev 3)}
Caveat!

Knot invariants cannot unambiguously classify a knot.

However

- knot invariants of the trivial knot and the first four knots are distinct from those of other prime knots with 10 crossings or fewer (249 knots in all), with one exception (5-1 and 10-132):

- Reidemeister moves and knot inflation can considerably reduce the number of possibilities.
Knot Inflation

Monte Carlo
Knot Tightening

Shrink-On-No-Overlaps (SONO) method of Piotr Pieranski. Scale all coordinates $s < 1$, keep bead radius fixed.
Results
Knotting Probabilities for Compact Lattice Loops
Chance of getting an unknot for several cube sizes:

\[ P_0(N) \approx e^{-N/N_0} \quad N_0 \sim 200 \]
Chance of getting the first few simple knots for different cube sizes:

![Graph showing the chance of getting the first few simple knots for different cube sizes. The graph plots the fraction of knot against length. The types of knots include trefoil (3-1), figure-eight (4-1), star (5-1), and (5-2). The graph indicates the probability peaks for each type of knot at different length intervals.]
Subchain statistics
Fragments of trivial knots are more *crumpled* compared to fragments of all knots.
Noncompact, Unrestricted Loop

Average gyration radius (squared) versus length

Closed random walk with fixed step length

Trivial knots *swell* compared to all knots for noncompact chains. This *topologically-driven* swelling is the same as that driven by self-avoidance (Flory exponent 3/5 versus gaussian exponent 1/2).

\[ R \approx N^{\frac{3}{5}} \]
Compact Lattice Loops
Ratios of average sub-chain sizes, trivial/all knots

Fragments of trivial knots are consistently more compact compared to fragments of all knots.
Compact Lattice Loops

General scaling of subchains (mean-square end-to-end) versus length

Over all knots: \( R \approx t^{\frac{1}{2}} \)

i.e. Gaussian;
Flory’s result for chains in a polymer melt.

Trivial knots: \( R \approx t^{\frac{1}{3}} \) ?

(A. Borovinskiy)
Knot (De)Localization
Localized or delocalized?
What have been shown computationally…

<table>
<thead>
<tr>
<th></th>
<th>Random circular chains</th>
<th>Flat knots (Polymer on sticky surface)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Noncompact</strong> (swollen)</td>
<td><strong>Localized</strong>*</td>
<td><strong>Localized</strong>**</td>
</tr>
<tr>
<td><strong>Compact</strong> (collapsed)</td>
<td>?</td>
<td><strong>Delocalized</strong>**</td>
</tr>
</tbody>
</table>

*Katritch, Olson, Vologodskii, Dubochet, Stasiak (2000). Preferred size of ‘core’ of trefoil knot is 7 segments.

**Orlandini, Stella, Vanderzande (2003). Localization to delocalization transition below a θ–point temperature.*
Knot Renormalization

Localized trefoil

$g=1$  $g=2$
Renormalization trajectory space
Renormalization trajectory
Initial state: Noncompact loop, N=384
Renormalization trajectory
Initial state: 8x8x8 compact lattice loop
Renormalization trajectory
Initial state: 12x12x12 compact lattice loop

![Graph of fraction of trefoil knots and trivial knots against fraction of trivial knots for different values of g.]
Knots in Proteins
Previous work…

Approx. 400 proteins, with random bridging of terminals, using Alexander polynomial. Found at most 3 knots.

3440 proteins, fixing the terminals and smoothing (shrinking) the segments in between. Found 6 trefoils and 2 figure-eights.

(Not about proteins) A study of linear random knots and their scaling behaviour.
Steps

1. Obtain protein structural information (.pdb files) from the Protein Data Bank. 4716 id’s of representative protein chains obtained from the Parallel Protein Information Analysis (PAPIA) system’s Representative Protein Chains from PDB (PDB-REPRDB).

2. Extract coordinates of protein backbone

3. Close the knot (3 ways)

4. Calculate knot invariants/signatures
Protein gyration radius versus length

\[ R = aN^{\frac{1}{3}} \]
CM-to-terminals distance versus gyration radius

\[ \langle C - T \rangle \approx 1.5R \]
DIRECT closure method

T1, T2 – protein terminals
CM-AYG closure method

C – center of mass
S1, S2 - located on surface of sphere surrounding the protein
F - point at some large distance away from C
RANDOM2 closure method

Study statistics of knot closures after generating 1000 pairs of points S1 and S2. Determine the dominant knot-type.
Knot probabilities in RANDOM2 closures for protein 1ejg chain A

N=46
Knot probabilities in RANDOM2 closures for protein 1xd3 chain A

N=229
Knot counts in the three closure methods

- **RANDOM2** and CM-AYG methods gave the same predictions for 4711 chains (out of 4716).
- **RANDOM2** and DIRECT methods gave the same predictions for 4528 chains (out of 4716).

<table>
<thead>
<tr>
<th>Closure method</th>
<th>Knot-type</th>
<th>0₁</th>
<th>3₁</th>
<th>4₁</th>
<th>5₁</th>
<th>5₂</th>
<th>6₁,up</th>
</tr>
</thead>
<tbody>
<tr>
<td>RANDOM2</td>
<td></td>
<td>4697</td>
<td>15</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>CM-AYG</td>
<td></td>
<td>4692</td>
<td>20</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>DIRECT</td>
<td></td>
<td>4516</td>
<td>164</td>
<td>9</td>
<td>9</td>
<td>3</td>
<td>15</td>
</tr>
</tbody>
</table>
Distribution of the % of RANDOM2 closures giving the dominant knot-type
Unknotting probabilities versus length for proteins and for compact lattice loops

Total of 19 non-trivial knots in the RANDOM2 method.
Knots in proteins occur much less often than in compact lattice loops.
Summary of Results

• Unknotting probability drops exponentially with chain length.  \( P_{\text{trivial}} \approx e^{-N/196} \)

• For compact conformations, subchains of trivial knots are consistently smaller than subchains of non-trivial knots. For noncompact conformations, the opposite is observed. *The fragments seem to be ‘aware’ of the knottedness of the whole thing.* (AYG)

• Knots in proteins are rare.
Unresolved issues…

• Are knots in compact loops delocalized? To what degree?
• Theoretical treatment of the scaling of subchains in compact loops with trivial knots.
• Theoretical prediction for the characteristic length of knotting $N_0$. $P_0(N) \approx e^{-N/N_0}$
Acknowledgments

A. Borovinskiy, N. Moore.
P. Pieranski and associates for
SONO animation.
DDF support, UMN Graduate School
Knot Mathematicians.
Biologists, Chemists and other researchers for
making protein structures available.