## Knotfind

We have extended the knotfind algorithm in this project to detect slipknots and visualize both knots and slipknots. Knotfind is an efficient knot prediction algorithm that analyzes every residue in a protein structure and searches for knots in them. The chain is continuously simplified by removing residues until only the termini remains. Knotfind then returns a simplified PDB file that contains only the residues which form the simplified chain.

**.1 Knot detection using Knotfind**

Knotfind uses only the alpha carbon atoms (cα) in a protein chain to detect the knots. The PDB is parsed into a Java method to extract only the data of the alpha carbon atoms and is stored onto a list. We only need the atom number and co-ordinates of the each cα atom. The atom data is later passed onto different methods for further processing of the chain.

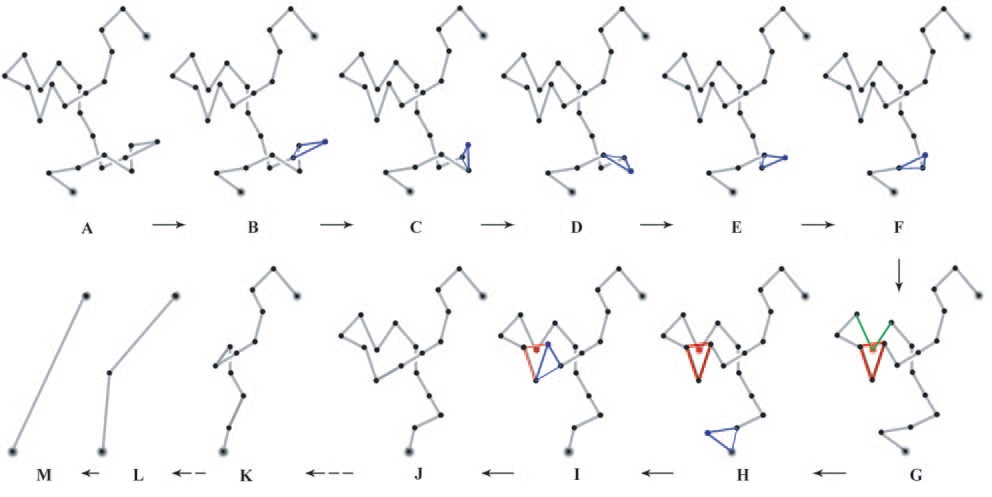
Knotfind uses an iterative approach to simplify and eliminate atoms from the residue chains. It initially starts with ‘n’ atoms (total number of cα’s in the given protein chain) and goes on until the size of the chain is two (for unknotted proteins) and proteins with knots will have two plus the unsimplified atoms.

Cα atoms arranged in the increasing order of i-1 to i+1 cartesian distance. Sets of three consecutive cα atoms i-1, i, i+1 are considered in each iteration. If there is no line segment, j, j+1 cutting through the triangle formed by connecting i-1, i, i+1, then cα i, is removed from the residue chain. If the line segment defined by j, j+1 is cutting through the triangle defined by i-1, i, i+1 then the cα i, is not simplified and the next set of i-1, i, i+1 is considered for simplification. This procedure is repeated until the last set of atoms are selected and simplified.

When the algorithm terminates it should only have N and C terminal cα atoms such that the chain has been simplified into a straight line. If the chain is not fully simplified, as in, there are other cα atoms remaining with the N and C terminals, and then those atoms define the knotted region of the protein1.

When a knot is detected, to double check and verify the knot an alternate method is also used, Where the area of the triangle formed by connecting i-1, i, i+1 is considered, If area covered by the triangle is being interested by any line j, j+1 which or also in the same plane then the i’th is not simplified else, the i’th atom is simplified and the process is repeated until the all the residues are simplified or checked. A tolerance of 0.0003 Å is used to round off errors. This is considered as a possible line width of the line connecting j, j+1[[1]](#endnote-1).

The algorithm keeps a log of all the simplified residues and the unsimplified residues. When the algorithm terminates the remaining residues are stored back onto a new PDB file. This is then visualized.



**Figure .1:** Simplification using Knotfind1.

In the above image1, we can see the changes in the state of the protein chain when the algorithm is running. Image A is the original structure of the backbone of the protein formed by connecting the cα atoms. The algorithm considers i-1, i, i+1such that the distances between them is the shortest. In image B, since no line segment is intersecting the lines connected by i-1, i, i+1 (The triangle marked in blue) the cα atom i is eliminated from the chain. The same process is repeated in the steps C through F. In the step G, an atom is intersecting the lines connected (The triangle marked in red), so i’th atom here is ignored and the next set i-1, i, i+1is considered for simplification. After iteration in steps H, one of the atoms with intersects in step G gets simplified and all the following atoms will get simplified in the later steps. When the algorithm terminates, since no knot is detected, it will have only the two terminal atoms left in the protein chain.

Knotfind algorithm can be used to benefit many structure prediction approaches, especially screening final models in a set of decoys and avoid a knotted model. Such screening methods are particularly important in automated methods like Robetta, where an expert will have to go in to check the final predictions manually. The speed of Knotfind makes it suitable not only for filtering of decoys to eliminate knotted structures but also during the protein structure prediction process as a filter or a part of the scoring scheme for optimization. Introduction of chain breaks in the modeling process is increasing the probability of knot formation. The location of chain breaks and the size of the gap generated by the chain breaks are important factors as well for the formation of knot1. To conclude, Knotfind algorithm is not only applicable to the homology based methods, but in any protein modeling process that introduce chain breaks during, including de novo prediction methods that have recently proved to have accuracy better than one second for small proteins with very few amino acids[[2]](#endnote-2).

1. Khatib, F.,Weirauch, M. T. & Rohl, C. A. (2006). Rapid knot detection and application to protein structure prediction. Bioinformatics. [↑](#endnote-ref-1)
2. Bradley,P. et al. (2005) Toward high-resolution de novo structure prediction for small proteins. Science [↑](#endnote-ref-2)