**KnotProt**

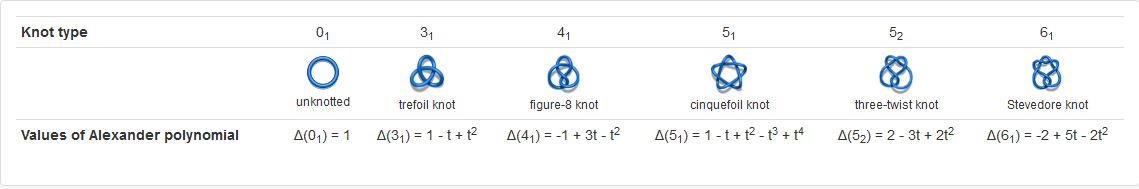
**.1 Introduction**

The KnotProt database collects information about topologically non-trivial proteins, i.e. proteins with knots and slipknots and represents them in the form of a “knotting fingerprint”[[1]](#endnote-1), and presents many statistics based on the obtained results. It is based on the proteins deposited in the protein data bank and has a database of 900 proteins with knots and slipknots[[2]](#endnote-2). Proteins form knots and slipknots in which backbone as a whole is unknotted1.

The details about every entanglement in a protein are stored in the KnotProt database and are presented in the form of a knotting fingerprint. The knotting fingerprint has the information about the type of the knot in each sub chain of the protein and represents it with a matrix diagram. The KnotProt database also provides extensive statistics about proteins with knots and slipknots based on their biological features and geometrical data such as depth of the knot and type of fold, length of the knot etc. KnotProt analysis reveals proteins having slipknots and knots can be classified into distinct topological motifs, represented by a few patterns in the matrix. This data can be used to find proteins with slipknots or knots based on the homological sequence, a similar structure, or something with particular biological function. The KnotProt website is updated after new proteins are deposited to the PDB every Wednesday.

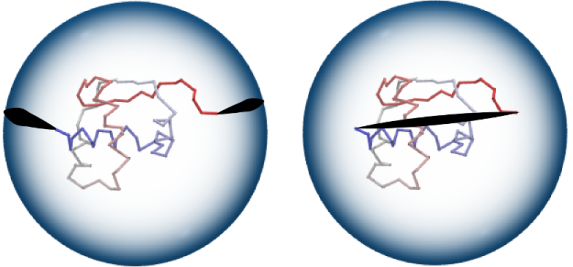
**.2 Knot Detection**

Many different types of knots have been found in proteins so far. A few of them are denoted as follows: trefoil knot (31), figure-8 knot (41), (52) and Stevedore’s knot (61). Unknotted loops are known as a trivial knot, or unknot and are denoted as 01. Knots are uniquely defined in closed chains. To define them in open chains such as proteins (which have loose ends), we choose how to connect the loose ends such that a loop is formed[[3]](#endnote-3). Making this choice optimally is a difficulty which has to be overcome while analyzing proteins. Once this choice has been made and once an open chain is transformed into a loop, the type of knot can be detected by calculating a knot invariant polynomial knows as Alexander polynomial[[4]](#endnote-4). It can be calculated by planar projection (two dimensional) of a knot. Alexander polynomial is different for all the knots with eight or fewer crossings, which is enough to detect knots in protein chains (The most complicated knot found in proteins so far has only six crossings).



**Figure .1:** Different types of knots and their associated Alexander Polynomials4.

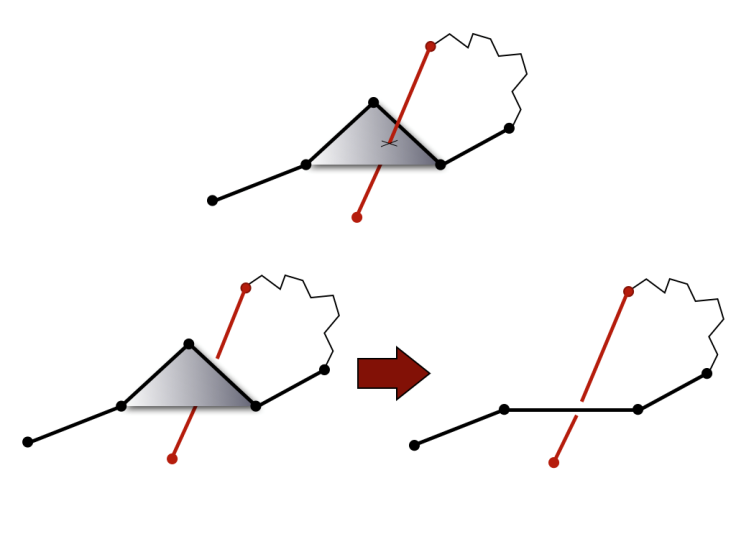
Defining knots in an open chain is nontrivial3. Knots can be uniquely determined only if the given protein chain is closed, and the classification of the knot depends on how one chooses to close the open chain[[5]](#endnote-5). KnotProt uses random closure method to close the open chains i.e. protein endpoints are connected several times to two randomly chosen points from a set of vertices of a truncated icosahedron placed on a sphere enclosing the given chain and these two points are connected by an arc on the circumference of the sphere. The knot associated with the given chain is then analyzed.



**Figure .2:** Random closure method (left). It is possible to choose a direct closure method (right), which connects end points based on the shortest interval.

The knot types that are obtained by connecting the open ended chains are determined by calculating the polynomial knot invariants. For quick calculations the Alexander polynomials are used.

Calculating knot polynomials for short protein chains is fast, but it might consume significant amount of time to compute long chains (ex: for protein chains with more than 500 amino acids). Therefore before computing the Alexander polynomial the length of the chain is reduced by applying the KMT algorithm[[6]](#endnote-6). The algorithm processes all the triangles formed by three consecutive amino acids and removes the middle atoms if there the no line segment cutting through it. After completing a number of iterations only the core of the protein which constitutes the knot remains and all other amino acids are removed. A similar approach is later explained while describing the knotfind algorithm.



**Figure .3:** Illustration of the KMT algorithm.

The KnotProt database not only checks if the chain is knotted but also analyzes all the subchains of given protein. For any protein this information is presented as the Knotting data with a matrix diagram[[7]](#endnote-7).

1. Sulkowska JI, Rawdon EJ, Millett KC, Onuchic JN and Stasiak A (2012) Conservation of complex knotting and slipknotting patterns in proteins. Proc. Natl. Acad. Sci. U.S.A. 109, E1715–E1723 [↑](#endnote-ref-1)
2. http://knotprot.cent.uw.edu.pl/introduction [↑](#endnote-ref-2)
3. Taylor WR (2000) A deeply knotted protein structure and how it might fold. Nature 406:916–919. [↑](#endnote-ref-3)
4. http://knotprot.cent.uw.edu.pl/knot\_detection [↑](#endnote-ref-4)
5. Millett, K.C., Rawdon, E.J., Stasiak, A. and Sułkowska, J.I. (2012) Identifying knots in proteins. Biochem. Soc. Trans. 41, 533–537. [↑](#endnote-ref-5)
6. Koniaris K, Muthukumar M (1991) Self-entanglement in ring polymers. J Chem Phys 95:2873–2881. [↑](#endnote-ref-6)
7. http://knotprot.cent.uw.edu.pl/knot\_plot [↑](#endnote-ref-7)