



Knot theory for proteins: Gauss codes, quandles and bondles

Colin Adams¹ · Judah Devadoss¹ · Mohamed Elhamdadi² · Alireza Mashaghi³ 

Received: 8 April 2020 / Accepted: 10 June 2020 / Published online: 1 July 2020
© The Author(s) 2020

Abstract

Proteins are linear molecular chains that often fold to function. The topology of folding is widely believed to define its properties and function, and knot theory has been applied to study protein structure and its implications. More than 97% of proteins are, however, classified as unknots when intra-chain interactions are ignored. This raises the question as to whether knot theory can be extended to include intra-chain interactions and thus be able to categorize topology of the proteins that are otherwise classified as unknotted. Here, we develop knot theory for folded linear molecular chains and apply it to proteins. For this purpose, proteins will be thought of as an embedding of a linear segment into three dimensions, with additional structure coming from self-bonding. We then project to a two-dimensional diagram and consider the basic rules of equivalence between two diagrams. We further consider the representation of projections of proteins using Gauss codes, or strings of numbers and letters, and how we can equate these codes with changes allowed in the diagrams. Finally, we explore the possibility of applying the algebraic structure of quandles to distinguish the topologies of proteins. Because of the presence of bonds, we extend the theory to define bondles, a type of quandle particularly adapted to distinguishing the topological types of proteins.

Keywords Topology · Knot · Protein · Gauss codes · Quandle · Bondle

✉ Alireza Mashaghi
a.mashaghi.tabari@lacdr.leidenuniv.nl

¹ Williams College, Williamstown, MA, USA

² University of South Florida, Tampa, FL, USA

³ Leiden University, Leiden, The Netherlands

1 Introduction

Folded linear molecular chains are ubiquitous in biology. Proteins and nucleic acids are linear polymers responsible for most cellular functions, for the inheritance of biological information, and are subject to changes during evolution and pathologies [8, 25]. These chains often fold to function and their 3D structure contains information about their dynamics, evolution, and inter-molecular interactions and can be used for designing drugs [15, 16]. Geometric and chemical properties of folded proteins and genomic DNA have been widely studied using various methods including NMR spectroscopy, X-ray crystallography, chromosome conformation capture, and mass spectrometry among others. Topological properties of these molecules have remained relatively unexplored due to lack of a relevant conceptual framework. Knot theory was successfully applied to study proteins and nucleic acids and protein and DNA knots were studied using various experimental techniques including nanopore technology and probe microscopy [24, 27, 29, 30]. Despite being interesting and innovative, these studies have had a limited impact on protein science as the vast majority of identified proteins fall into one topology class, i.e., the unknot [22].

Thus, standard knot theory cannot be effectively used to classify proteins [18]. Another shortcoming of the standard knot theoretic approach is that intra-molecular interactions or contacts are ignored. These interactions drive the folding of the molecular chains and are functionally important [11, 12, 17, 20, 26]. When intra-chain interaction is taken into consideration, the prevalence of knots and links substantially increases [5, 6, 28].

Thus, there is a need for a new topology framework that includes intra-chain interactions and is able to classify fold topology of biomolecular chains and in particular the proteins. This paper presents a new knot theory for folded linear molecular chains and looks to classify the topological structure of proteins through the application of certain aspects of knot theory. More specifically, we will apply a modified singular knot theory, Gauss codes to keep track of the structure, and an associated singular quandle called a bundle to distinguish structures.

Proteins are continuous linear molecules with the ends unbonded, so we will look at them as linear segments embedded in 3-space. Protein structure for a single protein is formed on three different levels, shown in Fig. 1. The Primary Structure is

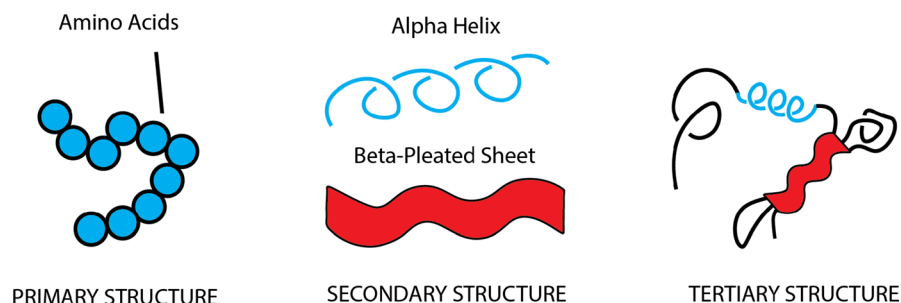


Fig. 1 Depiction of protein structures

defined by the ordering of the amino acids, or building blocks of the protein. These amino acids are bonded in a sequential chain from the N-terminus to the C-terminus, with each amino acid presenting an exposed R-group that can interact chemically or through the electrostatic effect with other R-groups and other molecules. The Secondary Structure is defined by the coiling or local structuring of the amino acids. This is where structural patterns appear such as β -pleated sheets and α -helices, both held together by hydrogen bonds. Tertiary Structure is defined by the interactions between different R-groups or backbone interactions, forming a structure for the entire protein.

A *knot* is an embedding of S^1 , the circle, into three dimensions, called a *conformation*, given by a function $f : [0, 2\pi] \rightarrow \mathbb{R}^3$ where $f(0) = f(2\pi)$. We define an equivalence on conformations of knots, considering two conformations equivalent if there is an ambient isotopy from one to the other. This means that we can deform the one through space to the other without passing the knot through itself. As is common, we will use the word “knot” to represent both a given conformation of a knot and the equivalence class of conformations corresponding to the given conformation, clarifying which is which when necessary.

Knot theory can be extended to singular knot theory, where we allow a finite number of singular points where two points on the circle are sent to the same point in 3-space by f . (See [7] for more on singular knots.) Allowing singularities will provide us with the ability to model intra-molecular bonds.

While it is often preferable to describe a knot in three dimensions, it is not always tractable. Therefore, we project the knot in a particular direction onto a plane, obtaining a *projection*. We consider only regular projections where a finite number of pairs of points on the knot are identified with each other and result in what we call crossings. We keep track of which of the two points in the pair is the top one. A projection is essentially a shadow that retains information at the crossings of the strands. Using a defined set of rules called Reidemeister moves, we can change one projection of a knot to any other projection of that same knot. In a two-dimensional projection, we define a *classical crossing* as a place where one strand goes over another, and if we are allowing singularities, a *singular crossing* where two or more strands intersect each other at a point in the three-dimensional conformation.

We can similarly think of proteins as a conformation of a line segment $[0,1]$ in three dimensions, which we will call a *protein model*, with an associated function $f : [0, 1] \rightarrow \mathbb{R}^3$. If we utilize the same equivalence of ambient isotopy, then every protein can be disentangled and they are all equivalent. But once we include singularities reflecting bonding of points along the conformation, this will no longer be true.

In a protein, we define singular crossings to exist where a protein has intra-chain interactions (also called contacts). These contacts take one of two forms. The first is covalent bonds, defined as two atoms sharing electrons. Of special interest to proteins are disulfide bridges, in which two thiol R-groups, made of one sulfur and one hydrogen, bond and release their hydrogen atoms. The second type of bond is formed through non-covalent interactions. A major form of non-covalent interactions are the electrostatic ones, particularly in the form of hydrogen bonds, which are partially electrostatic, but often come in multiples, making their strength significant

enough to control the protein's structure. Hydrogen bonds mediate the interactions between beta strands and the formation of alpha-helical structures.

Proteins differ from knots in that they have two endpoints that are not connected. One approach to this difference has been work on modelling proteins as *knotoids*, which are the projection of the conformations of a linear segment [6, 10]. Reidemeister moves (which we will look at in the next section) apply in the theory of knotoids, but in this theory, the endpoints cannot be moved across other strands in the projection. Although this may be useful when dealing with proteins that are somewhat rigid and have so-called lassos, here we desire a general theory that allows the movement of endpoints across strands in a projection. Thus we do not consider proteins as they relate to knotoids.

In finding a notation for protein structure, it is true that proteins have variable flexibility and restricted length, providing more limitations than the ones we place on curves in 3-space when defining knots. But the goal of this paper is topological in nature. Thus, we do not capture the full sense of rigidity or steric hindrance in a protein. Since we are allowing for the deformation of a protein's strands in the following sections, we will at times allow the same for the ends of the strand. The notation defined in this paper looks to balance simplicity and mathematical utility with chemical precision.

In Sect. 2, we discuss Reidemeister moves, which are moves one can do on a projection of a knot to obtain a new diagram of the same knot. We extend them to allow features present in proteins, including bonds, endpoint β -pleated sheets and α -helices.

In Sect. 3, we introduce Gauss codes, which can be used to describe in symbols a projection of a knot. We extend them to proteins.

In Sect. 4, we introduce quandles, which are algebraic objects that can be used to distinguish between different knots. In Sect. 5, we introduce singquandles, which are an extension of quandles that have been used to distinguish knots with singularities. We further extend this idea to the idea of a bundle, which is a quandle that can be applied in the presence of bonds.

In Sect. 6, we introduce the oriented bundle, which seems particularly suited to distinguishing between the topological types of proteins. We then identify several families of oriented bundles. In Sect. 7, we provide several examples of pairs of proteins that can be distinguished using oriented bundles.

2 Reidemeister moves

Reidemeister moves are a set of changes to the combinatorial pattern that is a projection of a knot. Planar isotopy is a deformation of the projection that does not change the combinatorial pattern. The critical result from [2] or [23] says that two knots K_1 and K_2 are equivalent if and only if there exists a sequence of Reidemeister moves and planar isotopy in the plane that transforms a projection of K_1 to a projection of K_2 . The three Reidemeister moves are depicted in Fig. 2. We refer to the monogonal face found in a Type I Reidemeister move as a *kink*.

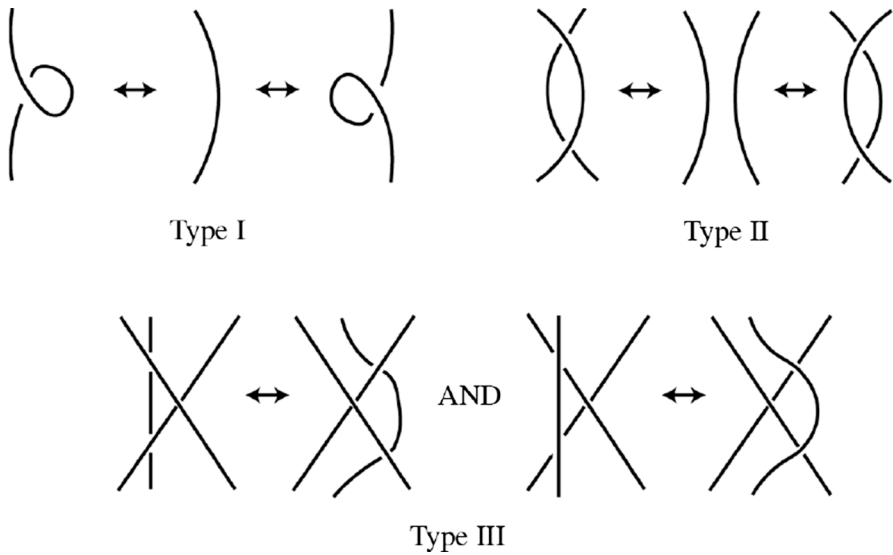


Fig. 2 Types I, II, and III Reidemeister moves

While these three Reidemeister moves are able to completely describe equivalence in classical knots, for our purposes we need to add in singular crossings to represent the covalent and hydrogen bonds that proteins form with themselves [7]. The use of singular crossings requires an additional two Reidemeister moves as shown in Fig. 3 [1, 31].

As appearing in this illustration, we denote a singularity by a small rectangle with two parallel edges on two strands. By drawing singular crossings in this fashion, we show that the strands do not cross over each other, but are instead bound together to more closely replicate the structure we see in proteins. By doing so, we do not lose any of the structure or properties we would hope to retain.

In a protein, a β -pleated sheet consists of multiple segments of a protein that run parallel to each other, roughly in a plane, with hydrogen bonds connecting each segment to its adjacent segment in multiple places. If we collapse the sheet to a point, this functionally looks like a singular crossing with more than two strands. Therefore, in order to represent β -pleated sheets, we must extend singular knot theory to contain multi-singularities. We define a *multi-singularity* as a



Fig. 3 Type IV and V Reidemeister moves for singularities

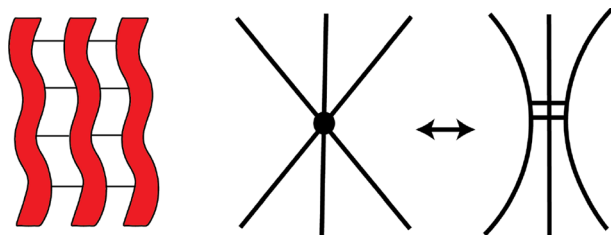


Fig. 4 β -pleated sheets and multi-singular crossings

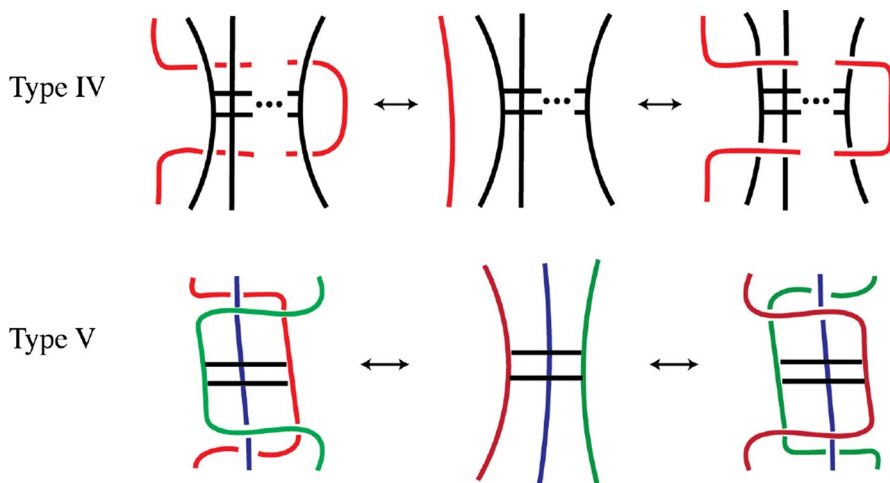


Fig. 5 Types IV and V Reidemeister moves for multi-singular crossings

place where two or more strands intersect each other at a single point, as shown in Fig. 4.

We can then incorporate this into the moves shown in Fig. 3 and define the set of singular Reidemeister moves shown in Fig. 5 to describe topological isotopy.

Note that when applying a Type V Reidemeister move to a multi-singular crossing, we allow any number of segments to be included, even though we have only depicted three segments to simplify the illustration. When the singularity is flipped, this causes a half-twist of all of the segments above and below the multi-singular crossing in the process.

Whereas in a Type IV move we only explicitly allow a segment to pass through a multi-singularity from left to right, we can use the given Reidemeister moves to show that we can slide a horizontal segment past a multi-singularity from top to bottom a well, as shown in Fig. 6.

Proteins also contain α -helices, which we can represent in multiple ways. An α -helix appears where a segment of a protein coils, with hydrogen bonds holding the coils together, as shown in Fig. 1. For now, we simply say that they do not add

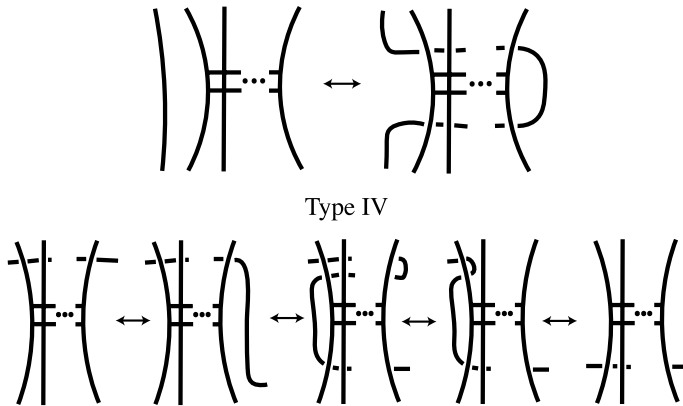


Fig. 6 A vertical Type IV Reidemeister move produces a horizontal Type IV Reidemeister move

topological structure, but are important in protein function. In a projection, we allow α -helices to slide along a segment, freely traversing a classical crossing, but not to pass through a singular crossing. Thus, we create a Reidemeister move Type VI to allow α -helices, depicted as the small set of jagged lines, to pass over or under other segments as in Fig. 7.

Since a protein is a continuous strand, we are able to equate its projection to a segment of a knot, allowing us the aforementioned Reidemeister moves. But because the protein has two unbonded endpoints, denoted N and C , that are free to move around, we need to be able to slide these endpoints past another strand. Therefore, we define a final Reidemeister Type VII move for this action as well as in Fig. 8.

We would like to prove that this set of seven moves captures all possible equivalences between projections of protein conformations. To achieve this, as is done

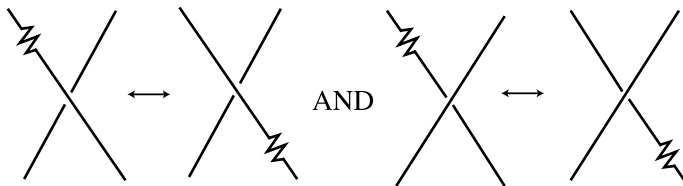


Fig. 7 Type VI Reidemeister move

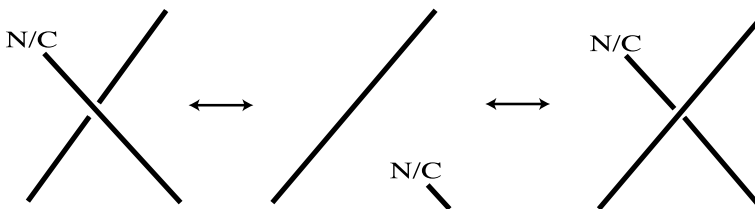


Fig. 8 Type VII Reidemeister move

in the knot case as well, we replace smooth conformations with piece-wise linear conformations, meaning that our conformation can be represented by a finite number of line segments glued end-to-end.

Theorem 2.1 *Given two protein conformations, they are equivalent if and only if there is a sequence of planar isotopies and Reidemeister moves from this set of seven moves to get from a projection of the first to a projection of the second.*

Proof As previously mentioned, it has already been proved that for a classical knot, planar isotopy and Type I, Type II, and Type III moves suffice to represent ambient isotopy in a knot (see [14] for a readable proof). The proof uses *triangle moves* to realize ambient isotopy between two piecewise linear knots. A triangle move is realized by taking a solid triangle that intersects the knot in one or two edges and replacing those line segments on the knot by the non-intersecting edges on the triangle, as shown in Fig. 9. Any isotopy we attain from deforming the original conformation can be represented by triangle moves. When we project down to a projection, one can show the triangle moves appear as planar isotopy and Reidemeister moves.

A spatial graph is a conformation of a graph (consisting of a collection of edges sharing a collection of vertices as their endpoints) in 3-space. A rigid vertex graph further posits that adjacent edges coming into a vertex cannot twist about one another. The vertex can be flipped, intertwining the edges appropriately as in our Type V move. In [14], it is shown that two conformations of a rigid vertex spatial graph with all vertices of valency 4 (called an RV4 graph) are equivalent if and only if there is a sequence of planar isotopies and Reidemesister moves of Types I, II, III, IV, and V from a projection of one to a projection of the other. This proof also considers how triangle moves impact the projection.

Our situation for a protein conformation has three differences from this one. First, we allow our vertices to have an even number of edges that can be four or greater. However, this does not impact the proof as in [14] and it goes through exactly as before.

Second, we have α -helices. These can be treated as vertices of valency two, and then the same arguments as in [14] go through to generate the Type VI Reidemeister move.

Third, we have the N and C endpoints of the protein. But it is straightforward to see that a triangle move that projects to overlap an endpoint will simply generate the Type VII Reidemeister move.

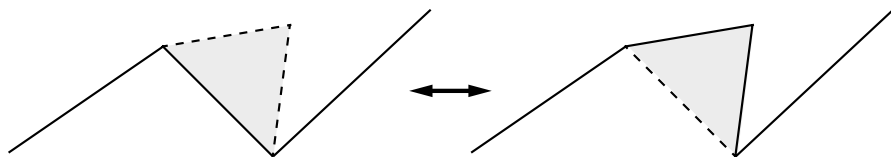


Fig. 9 Example of triangle move

Thus, all isotopies allowed in deforming our structure can be represented by triangle moves, which in the projections can be represented by the defined Reidemeister moves. Hence, the seven moves are sufficient to convert one projection of a protein conformation to any projection of a topologically equivalent protein conformation. \square

With these seven Reidemeister moves, we are able to transform one projection of a protein to another. However, it is inconvenient to convey these transformations through diagrams. Encoding diagrams as strings of text allows us to both interpret and transmit the information. Therefore, we turn to Gauss codes to represent protein structure in text form.

3 Gauss codes

Gauss codes are a means to represent knot projections using only symbols instead of diagrams. Using a Gauss code, we are able to easily go back and forth between the code and a projection while also being able to change entries in the code when Reidemeister moves are applied.

Since proteins are constructed and written from the N-terminus to the C-terminus, there is a definitive start and end point to proteins. Therefore, we start the Gauss code of a protein projection with N and end the code with C . We also have a natural orientation to the protein. Crossings are oriented as in Fig. 10.

As with traditional Gauss codes, when following a strand, if the strand crosses over another strand, we denote this with an O for over. Similarly, when traversing beneath another strand, we denote this with a U for under. We assign an orientation to the crossing denoted by a superscript of \pm .

If an α -helix appears, we denote it as α with a superscript of $+$ if the helix is right-handed (coils clockwise) and $-$ if the helix is left-handed (coils counterclockwise). Bonds are written as B , and β -pleated sheets are written as β . Strands in bonds and β -pleated sheets are labeled with a superscript of \pm , using $+$ if the strand runs parallel to the strand that first occurs in the bond or sheet (so the first strand always receives a $+$), and $-$ if the strand runs anti-parallel to the strand that first occurs.

With β -pleated sheets, strands are numbered with a subscript. The zero strand is defined as the first strand that appears in the protein's sequence. Numbers are assigned as sequential integers to the left and right of the initial strand, with positive integers appearing on the side to which the second strand appears in the

Fig. 10 Positive and negative oriented crossings

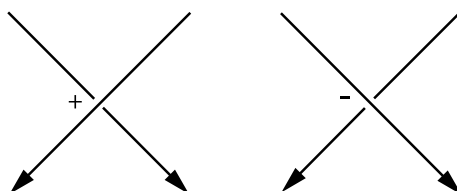


Fig. 11 A β -pleated sheet in a projection, with Gauss code $N \beta 1_0^+ \beta 2^+ \beta 1_2^- \beta 1_1^+ \beta 1_{-1}^-$

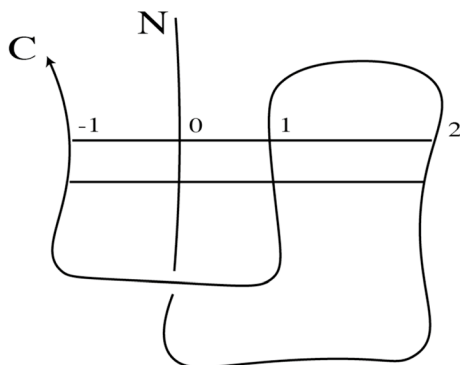
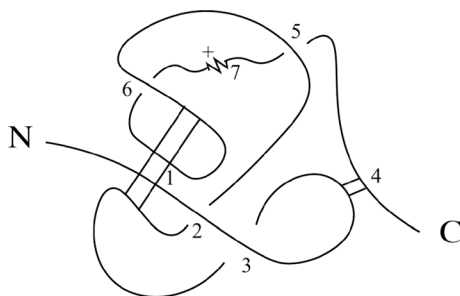


Fig. 12 Example of a projection of a protein conformation with Gauss code given by $N \beta 1_0^+ O 2^+ O 3^- B 4^+ U 3^- \beta 1_1^+ U 2^+ O 5^- O 6^+ \beta 1_{-2}^+ \beta 1_{-1}^- U 6^+ \alpha 7^+ U 5^- B 4^- C$



sequence, and all strands on the opposite side of the initial strand are defined as negative, as in Fig. 11.

Finally, each crossing, α -helix, bond, or β -pleated sheet is denoted with a sequential numbering based on its first appearance in the protein. Figure 12 gives an example of a Gauss code for a complete protein projection.

When we apply Reidemeister moves to a given projection, the move will be reflected in certain changes to the Gauss code. For instance, a Type I move inserts $On^\pm Un^\pm$ or $Un^\pm On^\pm$ into the Gauss code at the relevant point. Similar operations hold for all the Reidemeister moves.

For example, for the protein projection appearing in Fig. 12, we could apply a Type VI Reidemeister move to slide the α -helix out of the bigon (region in the projection plane bounded by two edges of the projection) bounded by crossings 5 and 6, and then remove the bigon by a Type II Reidemeister move to result in the Gauss code $N \beta 1_0^+ O 2^+ O 3^- B 4^+ U 3^- \beta 1_1^+ U 2^+ \beta 1_{-2}^+ \beta 1_{-1}^- \alpha 5^+ B 4^- C$.

But caution should be exercised. The corresponding operations on the Gauss codes do not always correspond to actual Reidemeister moves. For example, to do a Type II Reidemeister move, we must have two strands of the projection that are on the same complementary face of the projection. This is not visible from the Gauss code.

4 Quandles

A *knot invariant* is a map $I : \mathcal{K} \rightarrow S$ from all knot diagrams \mathcal{K} to some set S such that for any two projections K_1 and K_2 of the same knot type, $I(K_1) = I(K_2)$. The set S can be a collection of integers, groups, polynomials or other mathematical objects. An invariant is said to be a *complete invariant* if the converse is true, which is to say $I(K_1) = I(K_2)$ implies K_1 and K_2 represent the same knot type.

A quandle is an algebraic object that was introduced as an invariant for knots in 1982 independently in [13, 19]. It has turned out to be a particularly effective means to distinguish knots. For more details on quandles, see for example [9].

Definition 4.1 A *quandle* is a set X with an operation $\triangleright : X \times X \rightarrow X$ such that the following three conditions are satisfied.

- (1) For all $x \in X$, $x \triangleright x = x$.
- (2) There exists an inverse function \triangleright^{-1} such that for all $x, y \in X$,
 $(x \triangleright y) \triangleright^{-1} y = x = (x \triangleright^{-1} y) \triangleright y$.
- (3) For all $x, y, z \in X$, $(x \triangleright y) \triangleright z = (x \triangleright z) \triangleright (y \triangleright z)$.

A slightly more restrictive algebraic structure than a quandle is a *kei*, also called an *involutory quandle*.

Definition 4.2 A *kei*, or *involutory quandle* is a set X and operation $\triangleright : X \times X \rightarrow X$ that satisfy the following three conditions.

- (1) For all $x \in X$, $x \triangleright x = x$.
- (2) For all $x, y \in X$, $(x \triangleright y) \triangleright y = x = (x \triangleright y) \triangleright y$.
- (3) For all $x, y, z \in X$, $(x \triangleright y) \triangleright z = (x \triangleright z) \triangleright (y \triangleright z)$.

Note that the only difference is that for an involutory quandle, \triangleright is equivalent to \triangleright^{-1} .

Depending on the situation, as we will discuss, one or the other of these algebraic structures may be the more appropriate to apply.

A *coloring* of an oriented knot projection by a quandle is an assignment of a value from X to each arc, where an *arc* is defined as part of a strand in a projection that both starts and ends at an under crossing, but going over zero or as many crossings as we like. We require that the labels assigned to the arcs be related through the quandle operation as in Fig. 13.

The relevance of quandles to knots becomes apparent when we consider how the Reidemeister moves affect our labelled diagram as in Fig. 14. We see that the quandle axioms satisfied by the labels ensure that the quandle coloring is still valid after the Reidemeister moves. This means that the validity of the quandle coloring does not depend on the particular projection. It just depends on the knot type.

Fig. 13 Quandle conditions that must be satisfied at a crossing

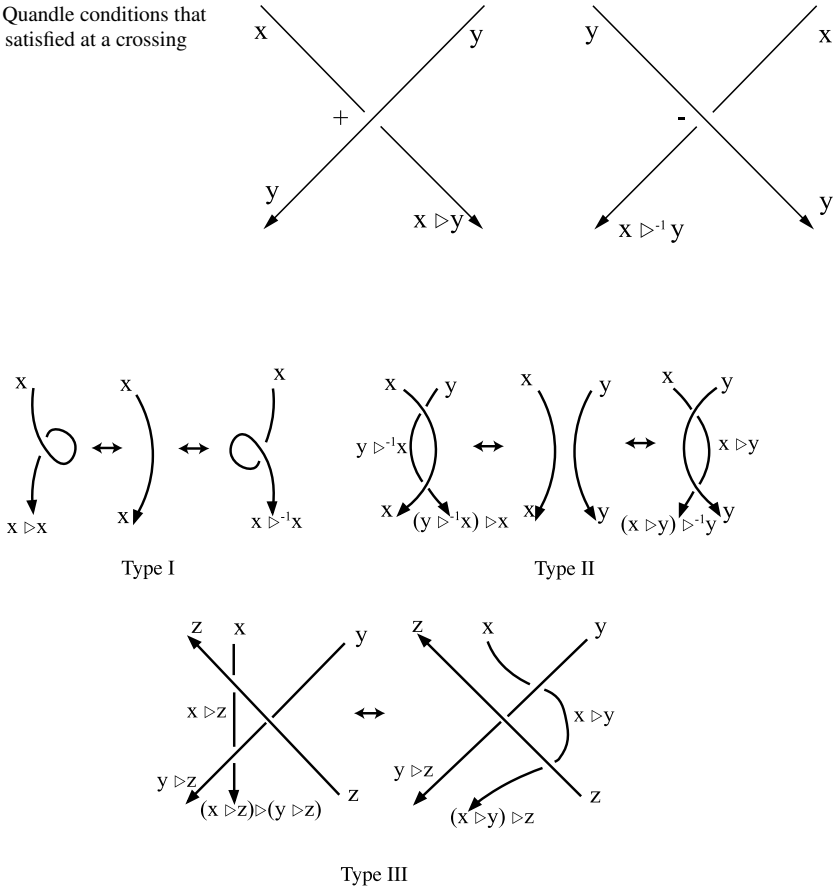


Fig. 14 The quandle relations guarantee the Reidemeister moves respect the labels

Thus, given a particular quandle, we can generate an invariant for knots by seeing how many distinct colorings by that quandle a particular knot has. Two knots with different numbers of colorings by that quandle must then be distinct knots.

We can also drop the orientation on the knots, in which case \triangleright and \triangleright^{-1} become identical, the arrows disappear in Fig. 14, and we color with involutory quandles instead. This simplifies things as we only have one operation to consider instead of two.

5 Quandles and singularities

In order to allow for singularities in knots, the authors of [4] introduced the *singquandle*. We first consider the singquandle for an unoriented knot, which will be an involutory quandle that satisfies additional conditions.

An arc in a singular knot projection is a strand that begins and ends at either an under-crossing or a singularity. Given an involutory quandle coloring of the arcs of a projection, we require the labels to satisfy conditions at the singular crossings as in Fig. 15, where $R_1(x, y)$ and $R_2(x, y)$ are maps from $X \times X$ to X yet to be specified.

Since the diagram in Fig. 15 can be rotated by 90° , 180° and 270° clockwise and the relation between the top pair of labels and the bottom pair of labels must be maintained, we immediately obtain certain relations that must be satisfied:

$$x = R_2(R_2(x, y), R_1(x, y)) \quad (\text{rotate } 180^\circ) \quad (1)$$

$$y = R_1(R_2(x, y), R_1(x, y)) \quad (\text{rotate } 180^\circ) \quad (2)$$

$$x = R_1(y, R_2(x, y)) \quad (\text{rotate } 270^\circ) \quad (3)$$

$$R_1(x, y) = R_2(y, R_2(x, y)) \quad (\text{rotate } 270^\circ) \quad (4)$$

$$y = R_2(R_1(x, y), x) \quad (\text{rotate } 90^\circ) \quad (5)$$

$$R_2(x, y) = R_1(R_1(x, y), x) \quad (\text{rotate } 90^\circ) \quad (6)$$

In [4], the authors show that in the presence of singularities, the only additional Reidemeister moves necessary are those coming from sliding a separate vertical strand on the left to the right behind or in front of a singularity, or flipping a singularity as in Fig. 16.

These moves generate the additional relations:

$$(y \triangleright z) \triangleright R_2(x, z) = (y \triangleright x) \triangleright R_1(x, z) \quad (7)$$

Fig. 15 Labels at a singularity

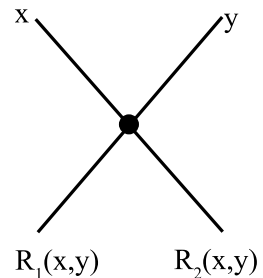
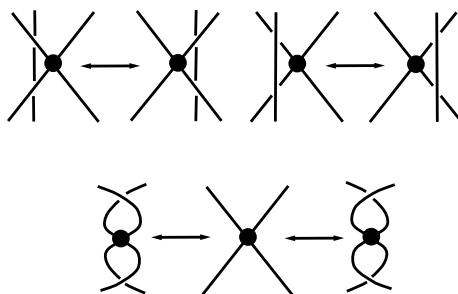


Fig. 16 Reidemeister moves for a singularity



$$R_1(x, y) = R_2(y \triangleright x, x) \quad (8)$$

$$R_2(x, y) = R_1(y \triangleright x, x) \triangleright R_2(y \triangleright x, x) \quad (9)$$

$$R_1(x \triangleright y, z) \triangleright y = R_1(x, z \triangleright y) \quad (10)$$

$$R_2(x \triangleright y, z) = R_2(x, z \triangleright y) \triangleright y \quad (11)$$

Definition 5.1 A *singquandle* is an involutory quandle, with a choice of $R_1(x, y)$ and $R_2(x, y)$ that satisfy all of the additional relations (1)–(11).

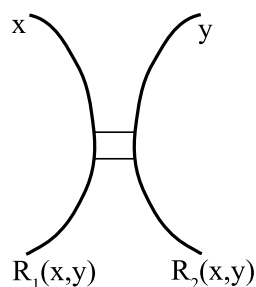
But the singularities we wish to consider for proteins are not of this type. In our case, we have bonds across two parallel strands, as in Fig. 17.

Such a bond does not have a four-fold rotational symmetry, but only a two-fold rotational symmetry. Thus, we have the following definition:

Definition 5.2 An *involutory bondle* is an involutory quandle that satisfies the relations (1), (2), (7), (8), (9), (10) and (11).

Although this choice allows us to incorporate bonds into our quandle, we do not yet have a way to represent β -pleated sheets. To deal with them, we replace a β -pleated sheet by a sequence of independent adjacent singular crossings as follows.

Fig. 17 Labels at a bond



We have already assigned positive and negative integer values to the strands in a β -pleated sheet from the subscripts of the Gauss code. Therefore, if we define the direction of the zero strand as downwards, we can define the relative heights of the individual singularities replacing a β -pleated sheet to be strictly increasing as the numbering increases, as shown in Fig. 18. The bonds appear as a set of stairs, either rising to the right or left, depending on which is the positive side of the labels on the β -pleated sheet. This transformation of a β -pleated sheet into adjacent singularities is called a *segmentation* of the β -pleated sheet.

With this, we can still perform the Type IV and Type V moves on multi-singularity crossing utilizing a sequence of Reidemeister moves on order two singularities, as shown in Fig. 19. Thus, no multi-singularity Reidemeister moves are needed. However, this choice for how to represent a β -pleated sheet does mean that we cannot distinguish between a protein with a β -pleated sheet and an identical one that has the corresponding sequence of bonds in place of the β -pleated sheet. Bondle invariants will be equivalent for the diagrams in Fig. 20.

The next issue we need to consider is the endpoints of the protein model. When considering proteins, we can view them as knot segments, with the ends free to move. Although in the physical realization of a protein, ends are sometimes tucked inside the protein or are subject to constraints and are therefore not free to move, in our model, we allow them to slide past strands in any given projection. Even for a fixed rigid conformation, as we change our projection direction, the endpoints in the projections can slide past strands, eliminating or creating crossings. Therefore, we treat the end strands as insignificant until they reach the first bond. We think of the ends as only being relevant in defining the first and last bonds, and ignoring them otherwise, as in Fig. 21.

The final structure we need to consider is the α -helix. We view it as a sequence of n kinks, where n is the number of full rotations that the helix contains, all having either + or – crossings depending on whether it is a clockwise or counterclockwise α -helix, as in Fig. 22. These kinks are referred to as *residues*. Just as we are unable to distinguish a segmented β -pleated sheet from a sequence of adjacent bonds, we cannot distinguish an α -helix from a sequence of kinks.

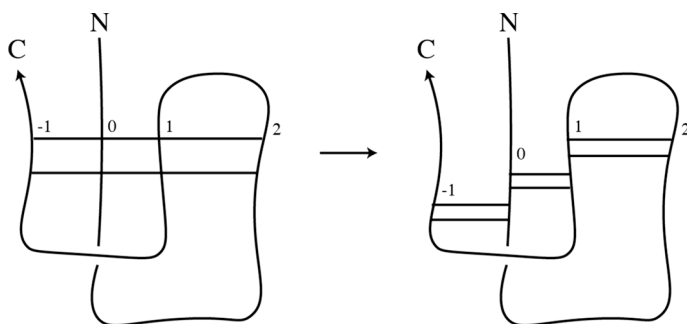


Fig. 18 A β -pleated sheet before and after segmentation

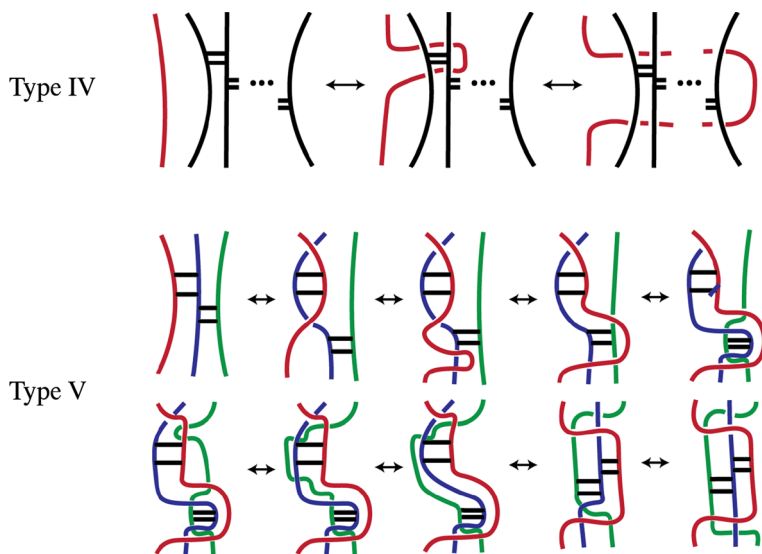


Fig. 19 Reidemeister moves on a segmented β -pleated sheet

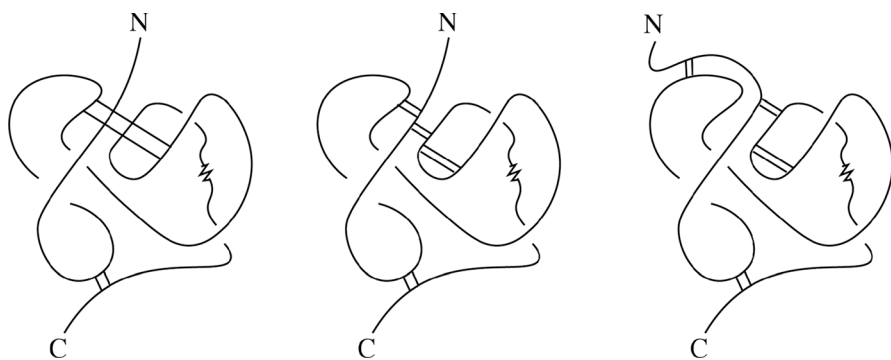


Fig. 20 Three indistinguishable protein models

When coloring a protein with a quandle, the α -helix becomes invisible because the Reidemeister Type I move in Fig. 14 allows for the removal of kinks. However, there is a generalization of a quandle called a rack that does not allow for the removal of kinks, and therefore does see the existence of an α -helix. A rack is simply a set that satisfies the second and third axioms of a quandle but not the first. We will not pursue racks further here.

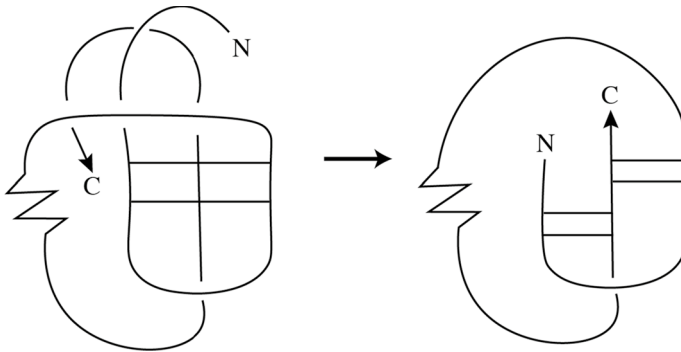


Fig. 21 Reducing end arcs

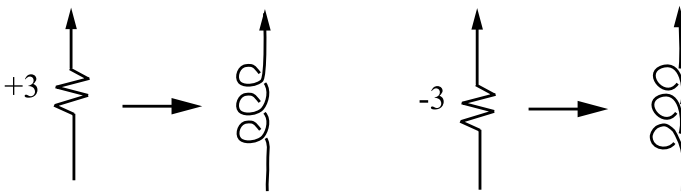


Fig. 22 Replacing an α -helix with a sequence of kinks

6 The oriented bundle

Since proteins do have a natural orientation, we should also consider the oriented version of the bundle. The oriented singquandle was defined in [3]. The authors showed that in addition to the four traditional Reidemeister moves on oriented diagrams that were shown in [21] to suffice for oriented links (appearing as the first four in Fig. 23), the 14 possible Reidemeister moves involving singularities for oriented links can be reduced to the three depicted in Fig. 23. Thus seven Reidemeister moves suffice for equivalency of singular diagrams.

Inserting our labels as in Fig. 15 (but with both strands pointing downward) into these three possibilities, we obtain a set of axioms to go with our three traditional quandle axioms coming from the non-singular moves.

Definition 6.1 Let (X, \triangleright) be a quandle. Then if R_1 and R_2 are two maps from $X \times X$ to X satisfying the following relations, we say that (X, \triangleright) is an *oriented singquandle*.

$$R_1(x \triangleright^{-1}y, z) \triangleright y = R_1(x, z \triangleright y) \tag{12}$$

$$R_2(x \triangleright^{-1}y, z) = R_2(x, z \triangleright y) \triangleright^{-1}y \tag{13}$$

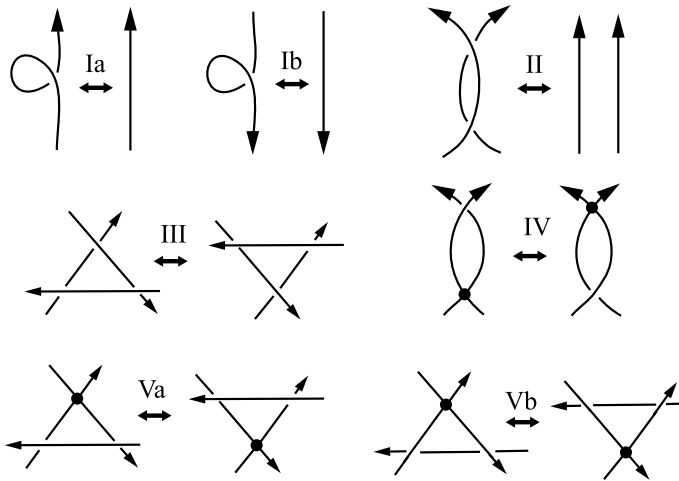


Fig. 23 A generating set of Reidemeister moves for oriented singular knots

$$(y \triangleright^{-1} R_1(x, z)) \triangleright x = (y \triangleright R_2(x, z)) \triangleright^{-1} z \tag{14}$$

$$R_2(x, y) = R_1(y, x \triangleright y) \tag{15}$$

$$R_1(x, y) \triangleright R_2(x, y) = R_2(y, x \triangleright y) \tag{16}$$

Note that for the oriented singquandle, there are no axioms coming from successive rotations by 90° of Fig. 15. The authors of [3] give the following two examples of singquandles.

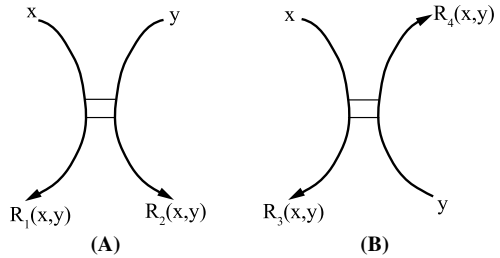
Example 6.2 Let n be a positive integer, and let a be an invertible element in \mathbb{Z}_n and b any element in \mathbb{Z}_n . Then the binary operations $x \triangleright y = ax + (1 - a)y$, $x \triangleright^{-1} y = a^{-1}x + (1 - a^{-1})y$, $R_1(x, y) = bx + (1 - b)y$ and $R_2(x, y) = a(1 - b)x + [b + (1 - b)(1 - a)]y$ make the triple $(\mathbb{Z}_n, \triangleright, R_1, R_2)$ satisfy the conditions to be an oriented singquandle.

Example 6.3 Let $X = G$ be a non-abelian group with the binary operation $x \triangleright y = y^{-1}xy$. Then, for $n \geq 1$, the following families of maps R_1 and R_2 make $(X, \triangleright, R_1, R_2)$ into an oriented singquandle:

1. $R_1(x, y) = x(xy^{-1})^n$ and $R_2(x, y) = y(x^{-1}y)^n$,
2. $R_1(x, y) = (xy^{-1})^n x$ and $R_2(x, y) = (x^{-1}y)^n y$,
3. $R_1(x, y) = x(yx^{-1})^{n+1}$ and $R_2(x, y) = x(y^{-1}x)^n$.

In the case of proteins, we would like to consider bonds rather than singularities. There are two distinct types of oriented bonds, one where the orientations on the two strands are parallel and one where they are anti-parallel, as in Fig. 24.

Fig. 24 Labels at bonds with parallel and anti-parallel strands



Each of the fourteen moves involving singularities from [3] yields two possibilities corresponding to whether the singularity is replaced with a vertical or horizontal bond. However, it is still true that for each of a vertical or horizontal bond, the fourteen moves reduce to three. So in addition to the four non-singular Reidemeister moves, we have six more moves to consider.

The first three correspond to bond diagram (A) in Fig. 24, and we inherit the same set of relations as for the singquandle, namely (12)–(16).

Considering bond diagram (B) from Fig. 24, we pick up two more functions $R_3(x, y)$ and $R_4(x, y)$. But note that rotation by 180° switches the roles of x and y and the roles of R_3 and R_4 . Thus, it is always the case that $R_4(x, y) = R_3(y, x)$. We will use this to eliminate $R_4(x, y)$ from all subsequent relations.

From Fig. 25, we obtain four additional relations.

Definition 6.4 An *oriented bundle* is a quandle with operation \triangleright and choices for functions $R_1(x, y), R_2(x, y)$ and $R_3(x, y)$ such that they satisfy relations (12)–(16) and the additional relations:

$$R_3(y, x \triangleright^{-1} z) = R_3(y \triangleright z, x) \triangleright^{-1} z \tag{17}$$

$$R_3(x, y \triangleright z) = R_3(x \triangleright^{-1} z, y) \triangleright z \tag{18}$$

$$(z \triangleright^{-1} R_3(x, y)) \triangleright x = (z \triangleright^{-1} y) \triangleright R_3(y, x) \tag{19}$$

$$R_3(x, y) \triangleright^{-1} y = R_3(x \triangleright^{-1} R_3(y, x), y). \tag{20}$$

Note that the maps $R_3(x, y) = x$ and $R_3(x, y) = y$ do always satisfy the relations (17), (18), (19) and (20) for any quandle (X, \triangleright) . We call these *trivial solutions* as they do not recognize the existence of the bond.

Since we already have examples of the desired maps R_1 and R_2 for both Example 6.2 and Example 6.3, we would like to find some solutions for the map R_3 satisfying relations (17), (18), (19) and (20).

Lemma 6.5 Let n be a positive odd integer greater than or equal to 3 and let a be an invertible element of \mathbb{Z}_n . Consider the quandle $(\mathbb{Z}_n, \triangleright)$ with $x \triangleright y = ax + (1 - a)y$ and inverse operation $x \triangleright^{-1} y = a^{-1}x + (1 - a^{-1})y$. Let m be an element in \mathbb{Z}_n and

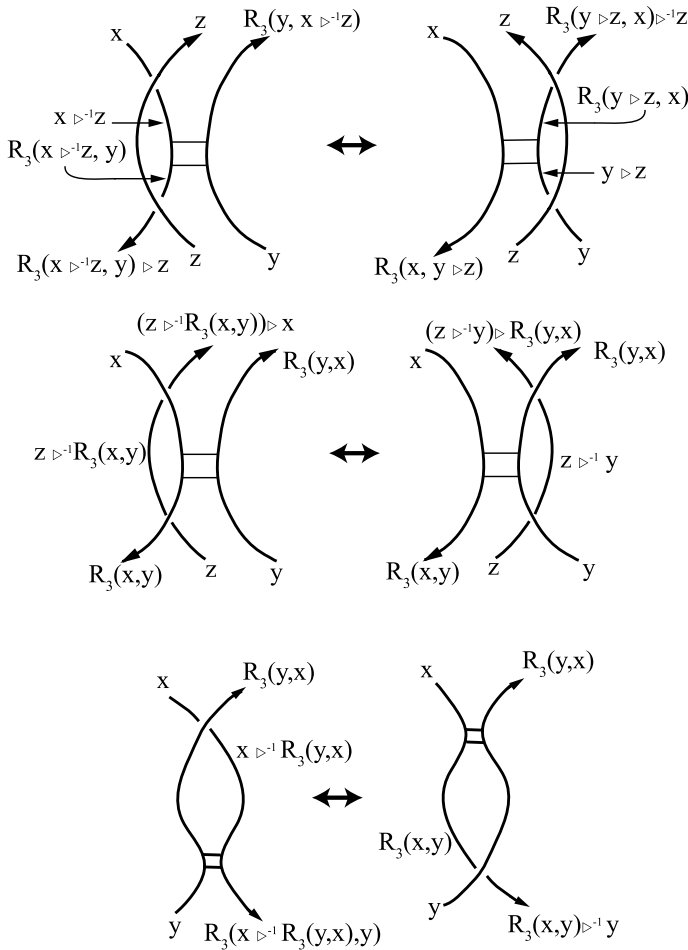


Fig. 25 Relations from oriented bonds

let R_3 be given by $R_3(x, y) = mx + (1 - m)y$. Then the map R_3 satisfies the Eqs. (17), (18), (19) and (20) if and only if $m(m - 1) = 0 \in \mathbb{Z}_n$.

Proof Direct computations show that the map R_3 given by $R_3(x, y) = mx + (1 - m)y$ satisfies the three Eqs. (17), (18), (19). Now substituting R_3 in Eq. (20) and simplifying gives the condition $m(m - 1)(x - y) = 0$, for all $x, y \in \mathbb{Z}_n$, and thus yields the condition $m(m - 1) = 0 \in \mathbb{Z}_n$. \square

We then have the following corollary

Corollary 6.6 Let $n = pq$ where p and q are odd primes. Assume further that $x \triangleright y = ax + (1 - a)y$ and $x \triangleright^{-1} y = a^{-1}x + (1 - a^{-1})y$, for

invertible element a in \mathbb{Z}_n . For any fixed element b in \mathbb{Z}_n , let $R_1(x, y) = bx + (1 - b)y$, $R_2(x, y) = a(1 - b)x + [b + (1 - a)(1 - b)]y$ and $R_3(x, y) = mx + (1 - m)y$. Then $(\mathbb{Z}_n, \triangleright, R_1, R_2, R_3)$ is an oriented bundle if and only if p divides m and q divides $(m - 1)$ or p divides $(m - 1)$ and q divides m .

The following is a list of some (n, m) satisfying Corollary 6.6.

1. If $n = 15$ then $m = 6$ or $m = 10$.
2. If $n = 21$ then $m = 7$ or $m = 15$.
3. If $n = 33$ then $m = 12$ or $m = 22$.
4. If $n = 35$ then $m = 15$ or $m = 21$.

Now we consider the case when the quandle is a group G with conjugation. First recall that the commutator of two elements x and y in a group G is given by $[x, y] := xyx^{-1}y^{-1}$. We have the following Lemma.

Lemma 6.7 *Let $X = G$ be a non-abelian group and let the quandle operation on G be given by $x \triangleright y = y^{-1}xy$, so that $x \triangleright^{-1}y = yxy^{-1}$. Assume that R_3 is given by $R_3(x, y) = x^p y^q$, where p and q are integers, then*

1. *The map R_3 satisfies both Eqs. (17) and (18) for any integers p and q .*
2. *If for all $x, y \in G$, $x^{p-1}y^q = x^{-q}y^{1-p}$ then R_3 satisfies Eq. (19).*
3. *Let p be an integer. If for all $x, y \in G$, the commutator $[x^p, y^{1-p}] = 1$, then R_3 satisfies Eq. (20).*

Proof Assume that R_3 has the form $R_3(x, y) = x^p y^q$, then

1. One can see that Eq. (17) is satisfied for all integers p and q from the following.

$$\begin{aligned} R_3(y, x \triangleright^{-1}z) &= y^p (zxz^{-1})^q = y^p zx^q z^{-1} = zz^{-1}y^p zx^q z^{-1} \\ &= R_3(y \triangleright z, x) \triangleright^{-1}z. \end{aligned}$$

Similarly, one has

$$R_3(x, y \triangleright z) = x^p z^{-1}y^q z = z^{-1}zx^p z^{-1}y^q z = R_3(x \triangleright^{-1}z, y) \triangleright z,$$

showing that Eq. (18) is satisfied also for all integers p and q .

2. Now we check Eq. (19). Assume that the equation $x^{p-1}y^q = x^{-q}y^{1-p}$ holds in G . Now we compute both the left hand side (LHS) and the right hand side (RHS) of Eq. (19).

$$\begin{aligned} LHS &= (z \triangleright^{-1}R_3(x, y)) \triangleright x = x^{-1} x^p y^q z y^{-q} x^{-p} x \\ &= x^{p-1}y^q z y^{-q}x^{1-p}, \end{aligned}$$

$$\begin{aligned} RHS &= [R_3(y, x)]^{-1}yz y^{-1}R_3(y, x) = (y^p x^q)^{-1}yz y^{-1}y^p x^q \\ &= x^{-q}y^{1-p} z y^{p-1}x^q. \end{aligned}$$

Since $x^{p-1}y^q = x^{-q}y^{1-p}$, then $LHS = RHS$ giving the result.

3. We finish by checking Eq. (20). Here also we compute separately the LHS and the RHS. We thus have

$$\begin{aligned} LHS &= R_3(x, y) \triangleright^{-1}y = yx^p y^q y^{-1} = yx^p y^{q-1} \\ RHS &= R_3(x \triangleright^{-1}R_3(y, x), y) = (y^p x^q x x^{-q} y^{-p})^p y^q \\ &= y^p x^p y^{-p} y^q = y^p x^p y^{q-p}. \end{aligned}$$

Now since the commutator $[x^p, y^{1-p}] = 1$, then we have $x^p y^{1-p} = y^{1-p} x^p$. Multiplying this equation by y^p from the left and by y^{q-1} from the right gives the equation $y^p x^p y^{q-p} = yx^p y^{q-1}$, thus we have $RHS = LHS$ giving Eq. (20). □

In order to give a more explicit example of a non-abelian group with a map R_3 satisfying Eqs. (19) and (20), we use the group of symmetries of a square.

Definition 6.8 Given a square with vertices labeled by 1, 2, 3 and 4, let G be the set of all rigid motions of the square that send vertices to vertices. Under composition, this set forms a non-abelian group called the *dihedral group of order 8* and denoted D_4 . Precisely, $D_4 = \{1, r, r^2, r^3, s, sr, sr^2, sr^3\}$, where the permutation $r = (1\ 2\ 3\ 4)$ is the clockwise rotation of 90° and s is the reflection $s = (1\ 2)(3\ 4)$.

Recall that in D_4 , the elements r and s satisfy the relations $r^4 = 1 = s^2$ and $srs = r^{-1}$. By iterating this last identity, we obtain $sr^i s = r^{-i}$, for $0 \leq i \leq 3$. The square of any element of D_4 is either the identity element 1 or r^2 . Then we see that r^2 commutes with any other element of D_4 , since $sr^i r^2 = (sr^{i+2} s) s = r^{-i-2} s = r^{4-i-2} s = r^{2-i} s = r^2 (sr^i s) s = r^2 sr^i$.

Corollary 6.9 In the dihedral group D_4 , the maps $R_3(x, y) = x^2 y^{-1}$ and $R_3(x, y) = x^{-1} y^2$ both satisfy Eqs. (19) and (20).

We thus obtain the following family of bundles.

Example 6.10 Let $X = D_4$ be the quandle with operation $x \triangleright y = y^{-1}xy$, then the following families of maps R_1, R_2 and R_3 make $(X, \triangleright, R_1, R_2, R_3)$ into a bundle:

1. $R_1(x, y) = x(xy^{-1})^n, R_2(x, y) = y(x^{-1}y)^n$ and $R_3(x, y) = x^2 y^{-1}$
2. $R_1(x, y) = (xy^{-1})^n x, R_2(x, y) = (x^{-1}y)^n y$ and $R_3(x, y) = x^2 y^{-1}$
3. $R_1(x, y) = x(yx^{-1})^{n+1}, R_2(x, y) = x(y^{-1}x)^n$ and $R_3(x, y) = x^2 y^{-1}$.

Note that this example still holds if we change $R_3(x, y) = x^2 y^{-1}$ to $R_3(x, y) = x^{-1} y^2$.

7 Examples

Given two projections of proteins and a choice of bundle, we can count the number of distinct colorings of each projection by that bundle, and if those numbers are distinct, we know the two proteins are not topologically equivalent. This provides an opportunity for the categorization of proteins into distinct topological types. In the following we give two examples demonstrating the use of *oriented bundles* to topologically distinguish proteins with bonds.

Example 7.1 In this example (Fig. 26), we use the oriented bundle $(\mathbb{Z}_{15}, \triangleright, R_1, R_2, R_3)$ from Corollary 6.6 with $a = 8$. Since 8×2 is congruent to 1 modulo 15 then $a^{-1} = 2$. We set $b = 2$ and then use this oriented bundle to distinguish the topological type of the following two two proteins P_1 and P_2 .

Precisely, $x \triangleright y = 8(x + y)$, $x \triangleright^{-1} y = 2x - y$, $R_1(x, y) = 2x - y$ and $R_2(x, y) = 7x - 6y$. Note that we do not need to define $R_3(x, y)$ because there are no anti-parallel bonds in the diagrams.

A coloring of P_1 gives the following equation

$$R_2(y, x) = y$$

which simplifies to

$$6(y - x) = 0$$

So if 5 divides $y - x$, we obtain a nontrivial coloring. Thus, the total number of colorings, including the trivial colorings, is 45.

On the other hand, a coloring of P_2 gives the following equation

$$R_2(y, x) \triangleright^{-1} R_1(y, x) = y$$

which simplifies to

$$11(y - x) = 0.$$

Since 11 is invertible in \mathbb{Z}_{15} , we see that $x = y$, implying that P_2 has *only* trivial colorings, of which there are 15. Thus P_1 and P_2 are distinct.

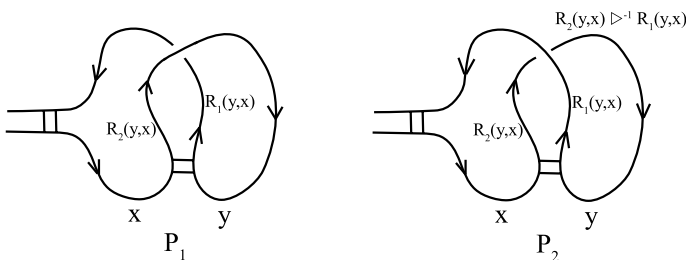


Fig. 26 Distinguishing P_1 from P_2

Example 7.2 In this example (Fig. 27), we include anti-parallel bonds. We utilize the bundle \mathbb{Z}_{15} with $a = 7$, $a^{-1} = 13 = -2$ (modulo 15), $b = 8$ and $m = 6$. Thus, $x \triangleright y = 7x - 6y$ and $x \triangleright^{-1} y = -2x + 3y$, $R_1(x, y) = 8x - 7y$, $R_2(x, y) = -4x + 5y$ and $R_3(x, y) = 6x - 5y$.

At the crossing with a hollow dot in P_1 , we obtain the relation:

$$y = [R_2(\alpha, R_3(x, y) \triangleright^{-1} R_1(\alpha, R_3(x, y))] \triangleright^{-1} R_3(y, x)$$

This yields $0 = 5(x - y)$, implying that there are nontrivial colorings corresponding to when 3 divides $x - y$. So we obtain a total of 75 colorings.

But at the crossing with a hollow dot in P_2 , we obtain the relation:

$$y = R_2(\beta, R_3(x, y) \triangleright \beta) \triangleright [R_1(\beta, R_3(x, y) \triangleright \beta) \triangleright R_2(\beta, R_3(x, y) \triangleright \beta)]$$

This yields $0 = 7(y - x)$, and as 7 is invertible, we only obtain the 15 trivial colorings corresponding to $y = x$. Thus, the two proteins must be topologically distinct.

8 Conclusion

When intra-chain interactions are included for linear molecules, a rich knot theory is possible. Utilizing some of the standard tools of knot theory extended to this new paradigm, including generalized Reidemeister moves and Gauss codes, it is possible to catalog the various knotted structures that result. To that end, the extension of quandles to bonded linear segments, called bundles, allows for the differentiation of the topological structures that can appear. This approach could be mechanized, allowing for computers to search for the parameters for the appropriate bundle to distinguish between the topological types of two proteins, for instance. There are many avenues for further research in these directions.

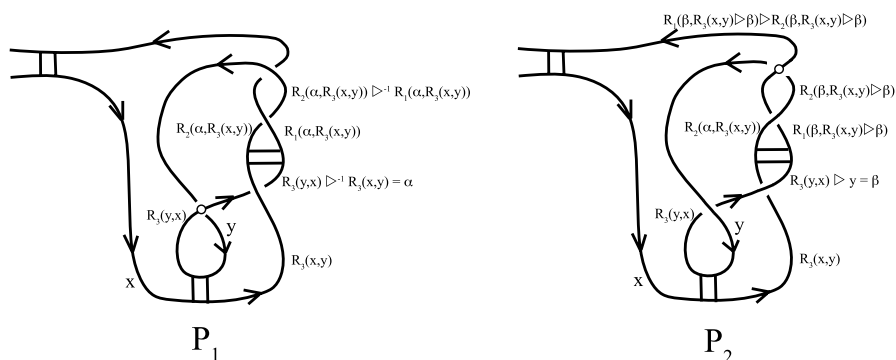


Fig. 27 Distinguishing P_1 from P_2

Acknowledgements Thanks to Jack Roche for suggesting the term “bundle”.

Funding Funding was provided by Muscular Dystrophy Association (USA), Grant Number MDA628071.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. E. Aceves, J. Elder, on invariants for spatial graphs. *Rose-Hulman Undergrad. Math. J.* **16**, 19 (2015)
2. J.W. Alexander, G.B. Briggs, On types of knotted curves. *Ann. Math.* **28**, 562–586 (1926)
3. K. Bataineh, M. Elhamdadi, M. Hajji, W. Youmans, Generating sets of Reidemeister moves of oriented singular links and quandles. *J. Knot Theory Ramif.* **27**(14), 1850064 (2018)
4. I.R.U. Churchill, M. Elhamdadi, M. Hajji, S. Nelson, Singular knots and involutive quandles. *J. Knot Theory Ramif.* **26**(14), 1750099 (2017)
5. P. Dabrowski-Tumanski, J. Sulkowska, Topological knots and links in proteins. *Proc. Natl. Acad. Sci.* **114**, 3145–3420 (2017)
6. P. Dabrowski-Tumanski, P. Rubach, D. Goundaroulis, J. Dorier, P. Sulkowski, K.C. Millett, E.J. Rawdon, A. Stasiak, J.I. Sulkowska, KnotProt 2.0: a database of proteins with knots and other entangled structures. *Nucleic Acids Res.* **47**(D1), D367–D375 (2019)
7. Z. Dancso, What is a singular knot? [arXiv:1811.08543 \[math\]](https://arxiv.org/abs/1811.08543) (2018)
8. C. Dobson, Protein-misfolding diseases: getting out of shape. *Nature* **418**, 729–730 (2002)
9. M. Elhamdadi, S. Nelson, *Quandles—An Introduction to the Algebra of Knots*, *Student Mathematical Library*, vol. 74 (American Mathematical Society, Providence, 2015)
10. D. Goundaroulis, N. Gügümcü, S. Lambropoulou, J. Dorier, A. Stasiak, L. Kauffman, Topological models for open-knotted protein chains using the concepts of knotoids and bonded knotoids. *Polymers* **9**, 444 (2017)
11. M. Heidari, V. Satarifard, S.J. Tans, M.R. Ejtehadi, S. Mashaghi, A. Mashaghi, Topology of internally constrained polymer chains. *Phys. Chem. Chem. Phys.* **19**(28), 18389–18393 (2017)
12. M. Heidari, V. Satarifard, A. Mashaghi, Mapping a single-molecule folding process onto a topological space. *Phys. Chem. Chem. Phys.* **21**(36), 20338–20345 (2019)
13. D. Joyce, A classifying invariant of knots: the knot quandle. *J. Pure Appl. Algebra* **23**, 37–65 (1982)
14. L.H. Kauffman, Invariants of graphs in three-space. *Trans. Am. Math. Soc.* **311**(2), 697–710 (1989). (en-US)
15. D.B. Kokh, J. Bomke, A. Wegener, H.P. Buchstaller, H.M. Eggenweiler, P.S. Matias, C. Wade, R.C. Amaral, M. Frech, Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nat. Commun.* **8**(2276), 1–14 (2017)
16. A.V. Luzhin, E.V. Nizovtseva, A. Safina, M.E. Valieva, A.K. Golov, A.K. Velichko, A.V. Lyubitelev, A.V. Feofanov, K.V. Gurova, V.M. Studitsky, O.L. Kantidze, S.V. Razin, The anti-cancer drugs curaxins target spatial genome organization. *Nat. Commun.* **10**(1441), 1–11 (2019)
17. A. Mashaghi, A. Ramezani, Circuit topology of linear polymers: a statistical mechanical treatment. *RSC Adv.* **5**(64), 51682–51689 (2015)
18. A. Mashaghi, R.J. van Wijk, S.J. Tans, Circuit topology of proteins and nucleic acids. *Structure* **22**(9), 1227–1237 (2014)
19. S. Matveev, Distributive groupoids in knot theory. *Math. USSR Sb.* **47**, 73–83 (1984)

20. A. Mugler, S.J. Tans, A. Mashaghi, Circuit topology of self-interacting chains: implications for folding and unfolding dynamics. *Phys. Chem. Chem. Phys.* **16**, 22537–22544 (2014)
21. M. Polyak, Minimal generating sets of Reidemeister moves. *Quantum Topol.* **1**, 399–411 (2010)
22. E.J. Rawdon, K.C. Millett, J.N. Onuchic, J.I. Sulkowska, A. Stasiak, Conservation of complex knotting and slipknotting patterns in proteins. *Proc. Natl. Acad. Sci. USA* **109**, E1715–E1723 (2012)
23. K. Reidemeister, Elementare begründung der knotentheorie. *Abh. Math. Sem. Univ. Hamburg* **5**, 24–32 (1927)
24. J.M. Rogers, A.L. Mallam, S.E. Jackson, Experimental detection of knotted conformations in denatured proteins. *Proc. Natl. Acad. Sci. USA* **107**, 8189–8194 (2010)
25. J. Rowley, V.G. Corces, Organizational principles of 3d genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018)
26. V. Satarifard, M. Heidari, S. Mashaghi, S.J. Tans, M.R. Ejtehadi, A. Mashaghi, Topology of polymer chains under nanoscale confinement. *Nanoscale* **9**(33), 12170–12177 (2017)
27. M.A. Soler, P.F. Faisca, Effects of knots on protein folding properties. *PLoS ONE* **8**, e74755 (2013)
28. A. Stasiak, J.I. Sulkowska, P. Dabrowski-Tumanski, D. Goundaroulis, θ -curves in proteins (2019). [arXiv:1908.05919](https://arxiv.org/abs/1908.05919)
29. P. Sulkowski, P. Szymczak, J.I. Sulkowska, M. Cieplak, Stabilizing effect of knots on proteins. *Proc. Natl. Acad. Sci. USA* **105**, 19714–19719 (2008)
30. W.R. Taylor, K. Lin, Protein knots: a tangled problem. *Nature* **421**, 25 (2003)
31. W. Yuasa, A_2 colored polynomials of rigid vertex graphs. *N. Y. J. Math.* **24**, 355–374 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.