# LOOPS: Visual Analysis Tool for Non-Periodic Protein Structures

Pavol Ulbrich[1,2], Michal Horňák[1], and Barbora Kozlíková[1]

[1]VisIt Lab, Masaryk University of Brno, Czech Republic
[2]VyGLab Research Laboratory, Universidad Nacional del Sur, Bahía Blanca, Argentina.

**Abstract**

*Protein molecules, the basic building blocks of all living organisms, are formed by chain sequences of amino acids. Due to internal forces, these chains are folded into a spatial representation with recurring patterns along the chain, called secondary structures. The basic secondary structures are α-helices and β-sheets, however, the protein function is also determined and influenced by other, non-periodic structural units, such as loops. Playing a pivotal role in the protein folding and dynamics, the loop classification and understanding may be necessary for modeling of shared structure ancestry (homology), protein structure prediction, and protein design. However, their non-repetitive characteristic denies the easy identification of their regularities. Therefore, we present LOOPS, a web-based tool for exploring and analyzing loops in proteins, that offers a high level of interaction with reference to other secondary structures. Displaying the additional information about a loop, such as its conformation, geometry, and multiple b-factors obtained from various methods, LOOPS provides a simple tool for analysis of loops in a protein sequence.*

**CCS Concepts**
• ***Human-computer interaction*** → *Biomedical visualization;*

## 1. Introduction

Understanding the behaviour of proteins is a challenging task, because it is effected by their structure, dynamic behavior, and other aspects. The basic building block of all proteins is the sequence of amino acids, known as a primary structure of a protein, followed by periodic secondary structures, such as α-helices and β-sheets, and the tertiary structure, representing folding of the whole sequence into a compact 3D shape. Recent studies suggest the protein behaviour is also influenced by non-periodic structures, known as loops [AIM08, PIBGG*13]. Loops are segments of amino acids joining the periodic secondary structures, which have significant influence on protein-protein interactions, ligand and DNA binding, enzyme catalysis, etc. [FS03]. So far, the loop classification has been practically applied to predictions of protein functions [LRO07] and protein interaction [MPC*13].

Despite the clear benefits of loop classification, there have not been any attempts to provide an interactive visual representation of these non-periodic structures, enabling to explore loop structures and their properties. Therefore, here we propose a visual tool aiming to display the loop positions and their properties. We believe that our approach might simplify and ease the process of understanding the importance of loops for protein function. We are designing our tool in tight collaboration with the domain experts, providing us also with the input data and valuable feedback.

## 2. Loop characteristic

Over the course of the past years, several methods for loop classification were presented [AHB*04, EFFH*04, BGGPI*13], defining a non-periodic structure as an elementary super-secondary motif—one loop plus its bracing secondary structures [SHOT99] (see Figure 1). In other words, a loop is defined as a subsequence of amino acids that starts with the first amino acid belonging to the preceding secondary structure and ends with the last amino acid of the closing secondary structure. Between these secondary structures lays a so-called *coil*, which is formed by amino acids, that do not belong to any secondary structure in the protein sequence.

**Figure 1:** *A protein loop is defined as a sequence of amino acids spanning a triplet of two periodic secondary structures and a coil in-between them.*

For the classification purposes in ArchDB [BGGPI*13], a loop



DOI: 10.2312/eurp.20191154

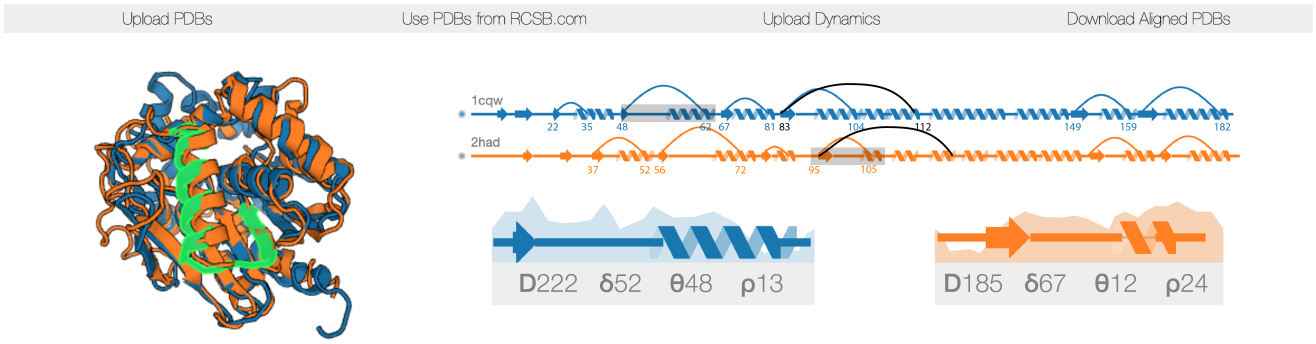


**Figure 2:** *The user interface of LOOPS. Three basic views combine 3D view, 1D sequential representation equipped with secondary structures and loops, and finally detailed views on selected loops with additional information such as loop geometry and b-factors.*

is defined by its length (number of amino acids), conformation ($\phi$ and $\psi$ backbone dihedral angles of the amino acids), the bracing secondary structures, and the geometry of the loop. The geometry is defined by four internal coordinates (D, $\delta$, $\theta$, $\rho$), extracted from the orientation of the principal vectors (M1, M2) that define the bracing secondary structures [OBQ*97].

Another important aspect in the loop analysis is the b-factor number. It represents the fluctuation of atoms around their average positions and provides us with essential information about protein dynamics [YBT05]. Such information can produce a valuable input to the loop definition process as we can observe the fluctuation of not only atoms but whole amino acids, secondary structures, and whole loops as well.

## 3. Design and Implementation

Addressing all the features and loop properties described above, we designed LOOPS, a visual tool for analysis of protein loops. The implementation extends the approach presented by Kocincová et al. [KJB*17], serving namely for comparison of protein secondary structures by encoding the spatial orientation of secondary structures directly into the 1D sequential view. The same composition of views and interaction can be used to compare the positions and properties of loops. Additionally, it provides robust layout for observation of one or more protein sequences.

The LOOPS application benefits from spatial alignment of the compared protein chains, which can be observed in the 3D view, and illustrates the spatial conformation of the loops. Their relative position in the chains can be compared in the juxtaposition view. Continuous values of b-factors are plotted in the area chart behind the detailed view of a selected loop, distinguishing protein parts with high fluctuation of atoms from regions with low one. Loop geometry values are displayed for easier categorisation and interpretation.

The LOOPS application consists of the following linked views. The first view contains the PV viewer [BBW*14] JavaScript module, which supports the 3D visualization of uploaded proteins and basic interactions, such as rotation, zoom, and selection. The uploaded protein structure is straightened into the second, 1D view. Here we convey the positions of secondary structures within the sequence as glyphs (spirals and arrows) and detected loops as arches above them. For every loop, the sequence number taken from the PDB file of the starting and ending amino acid is displayed. After a loop is selected, it is immediately highlighted in the PV viewer window and visualized also in the third proposed view, showing detailed information about a selected loop and its properties.

The detailed view shows a zoomed-in loop with additional information, such as its geometric properties and average b-factors of amino acids in the loop. The latter is visualized as a area chart behind the glyphs, instantly identifying amino acids within the loop with high b-factor fluctuation numbers. The position and geometry of the loop are obtained from the ArchDB loop database [BGGPI*13]. The 3D position of amino acids, secondary structures layout, and b-factors are accessed directly from the PDB file containing the protein.

All views are interactively linked—the selection of a loop in the 1D sequence view leads to changes in the 3D and the detailed loop view. This enables the user not only an intuitive exploration of a single protein structure and its loops, but our tool also addresses the request of the collaborating domain experts to be able to compare two or more protein sequences and their loops. Figure 2 shows the interplay between our proposed visualizations, used for comparison of two loops, each coming from one protein.

## 4. Conclusion and Future Work

In this paper, we presented the design and implementation of a prototype for the web-based LOOPS application. It is a web-based tool for visualization and interaction with loop structures in protein sequences. The right combination of views grants the ability to explore the properties of individual loops and compare them with other loops within one protein sequence or across more proteins. In the near future, the application will be tested on selected datasets from the ArchDB database. Additionally, it will be further developed to satisfy the demands of new trends in loop classification. This will include the support for so called superloops (suggested in Figure 2 as black arches) and dynamic data inputs, such as b-factors calculated *ex-novo* and percentage of structure over time.

## References


[AHB*04] ANDREEVA A., HOWORTH D., BRENNER S. E., HUBBARD T. J., CHOTHIA C., MURZIN A. G.: Scop database in 2004: refinements integrate structure and sequence family data. *Nucleic acids research 32*, suppl_1 (2004), D226–D229. 1

[AIM08] AKIVA E., ITZHAKI Z., MARGALIT H.: Built-in loops allow versatility in domain–domain interactions: Lessons from self-interacting domains. *Proceedings of the National Academy of Sciences 105*, 36 (2008), 13292–13297. doi:10.1073/pnas.0801207105. 1

[BBW*14] BIASINI M., BIENERT S., WATERHOUSE A., ARNOLD K., STUDER G., SCHMIDT T., KIEFER F., CASSARINO T. G., BERTONI M., BORDOLI L., SCHWEDE T.: SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research 42*, Webserver-Issue (2014), 252–258. 2

[BGGPI*13] BONET J., GARCIA-GARCIA J., PLANAS-IGLESIAS J., MARÍN-LÓPEZ M. A., OLIVA B., FERNANDEZ-FUENTES N.: ArchDB 2014: structural classification of loops in proteins. *Nucleic Acids Research 42*, D1 (2013), D315–D319. doi:10.1093/nar/gkt1189. 1, 2

[EFFH*04] ESPADALER J., FERNANDEZ-FUENTES N., HERMOSO A., QUEROL E., AVILES F. X., STERNBERG M. J., OLIVA B.: Archdb: automated protein loop classification as a tool for structural genomics. *Nucleic Acids Research 32*, suppl_1 (2004), D185–D188. 1

[FS03] FISER A., SALI A.: ModLoop: automated modeling of loops in protein structures. *Bioinformatics 19*, 18 (Dec 2003), 2500–2501. 1

[KJB*17] KOCINCOVÁ L., JAREŠOVÁ M., BYŠKA J., PARULEK J., HAUSER H., KOZLÍKOVÁ B.: Comparative visualization of protein secondary structures. *BMC Bioinformatics 18*, 2 (Feb 2017), 23. doi:10.1186/s12859-016-1449-z. 2

[LRO07] LEE D., REDFERN O., ORENGO C.: Predicting protein function from sequence and structure. *Nature Reviews. Molecular Cell Biology 8*, 12 (2007), 995. 1

[MPC*13] MOSCA R., PONS T., CÉOL A., VALENCIA A., ALOY P.: Towards a detailed atlas of protein–protein interactions. *Current Opinion in Structural Biology 23*, 6 (2013), 929–940. 1

[OBQ*97] OLIVA B., BATES P. A., QUEROL E., AVILÉS F. X., STERNBERG M. J.: An automated classification of the structure of protein loops. *Journal of Molecular Biology 266*, 4 (1997), 814–830. 2

[PIBGG*13] PLANAS-IGLESIAS J., BONET J., GARCÍA-GARCÍA J., MARÍN-LÓPEZ M. A., FELIU E., OLIVA B.: Understanding protein-protein interactions using local structural features. *Journal of Molecular Biology 425*, 7 (2013), 1210 – 1224. doi:https://doi.org/10.1016/j.jmb.2013.01.014. 1

[SHOT99] SALEM G. M., HUTCHINSON E. G., ORENGO C. A., THORNTON J. M.: Correlation of observed fold frequency with the occurrence of local structural motifs. *Journal of Molecular Biology 287*, 5 (1999), 969–981. 1

[YBT05] YUAN Z., BAILEY T. L., TEASDALE R. D.: Prediction of protein b-factor profiles. *Proteins: Structure, Function, and Bioinformatics 58*, 4 (2005), 905–912. 2