

---

## Sequence Analysis

# ProSeqViewer: an interactive, responsive and efficient TypeScript library for visualization of sequences and alignments in web applications

Martina Bevilacqua<sup>1</sup>, Lisanna Paladin<sup>2</sup>, Silvio C.E. Tosatto<sup>1,\*</sup>, Damiano Piovesan<sup>1,\*</sup>

<sup>1</sup>Department of Biomedical Sciences, University of Padua, Padova 35121, Italy.

<sup>2</sup>Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), 69117, Heidelberg, Germany.

\*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

### Abstract

**Summary:** Biological data is ever-increasing in amount and complexity. The mapping of this data to biological entities such as nucleotide and amino acid sequences supports biological data analysis, classification and prediction. Sequence alignments and comparison allow the transfer of knowledge to evolutionary related entities, the mapping of functional domains, the identification of binding and modification sites. To support these types of studies we developed ProSeqViewer, a tool to visualize annotation on single sequences and multiple sequence alignments. This state-of-the-art multifunctional library was developed as a modular component to be integrated into static or dynamic web resources and support intuitive visualization of sequence features. ProSeqViewer is extremely lightweight, fast, interactive, dynamic, responsive and works at any screen size. It generates pure HTML which is compatible with any browser and operating system. ProSeqViewer can exchange events with other visualization components and is already used by multiple biological databases.

**Availability and implementation:** ProSeqViewer is an open-source TypeScript library compatible with state-of-the-art website environments. The source code and an extensive documentation including use cases are available from the URL: <https://github.com/BioComputingUP/ProSeqViewer>.

**Contact:** [silvio.tosatto@unipd.it](mailto:silvio.tosatto@unipd.it); [damiano.piovesan@unipd.it](mailto:damiano.piovesan@unipd.it)

---

## 1 Introduction

Molecular sequences are by far the most abundant type of biological data available in public repositories. Multiple sequence alignments (MSAs) are particularly useful to study biological sequences by providing the starting point for example to infer evolutionary relationships and identify functional features. Two categories of tools are available to visualize annotations and key positions directly into the sequence, stand-alone software packages and web applications. Both serve different types of users and offer different functionality. Stand-alone software packages provide higher control and can be more powerful in managing big MSAs. On the other hand, web applications are easier to use, do not require programming skills and can benefit from the interoperability and integration of web data. Stand-alone packages for visualizing MSAs include ESPript (Gouet *et al.*, 2003), JalView (Waterhouse *et al.*, 2009, 2) and Mview (Brown *et al.*, 1998). ESPript is a FORTRAN program used to generate high-resolution MSA graphics for publications. The user needs to provide all annotations and configurations via commands in a custom scripting language. JalView is a general purpose Java

platform to visualize and manipulate MSAs. It offers a graphical interface and is available both as stand-alone and as browser plugin. MView is a Perl package that can generate HTML from a Fasta input. It is available both as a stand-alone and as a web service at the European Bioinformatics Institute (EBI).

Sequence and MSA viewers are key components of most modern core data resources, such as UniProtKB (The UniProt Consortium, 2021) and Pfam (Mistry *et al.*, 2021), where annotations and key positions are directly highlighted in the sequence. In most cases the visualization component of a web page is directly compiled in the website bundle. In other cases it is made available as an independent JavaScript library and can be reused by other web resources.

Reusable web libraries include the MSASviewer (Yachdav *et al.*, 2016), the NextProt sequence viewer (Zahn-Zabal *et al.*, 2020) and LexiconMonoSeq (Tanyalcin, 2019). They are not simple static viewers but can modify the view dynamically, based on a number of available commands. MSASviewer is tailored for MSA visualization, rendering a sequence logo and providing predefined color palettes for amino acid propensities. The NextProt sequence viewer is very simple and easy to integrate but designed for single sequences and not MSAs.

LexiconMonoSeq is optimized for the visualization of long sequences, e.g. DNA genomic fragments, featuring lazy asynchronous loading. It is not specific to biological sequences and thus misses typical visualization features such as physico-chemical color schemes, pattern searches and consensus visualization. None of these tools are responsive to window changes and all rely heavily on external dependencies slowing down the website loading.

Here we present ProSeqViewer, a TypeScript library for both single sequence and MSA visualization. The tool offers higher flexibility and rendering performance, captures mouse events and is natively responsive to changes in window size. The software is accompanied by extensive documentation including usage examples, demos and instructions about how to integrate the tool in modern web applications. ProSeqViewer is already employed in the DisProt (Hatos et al., 2020), MobiDB (Piovesan et al., 2021) and RepeatsDB (Paladin et al., 2021) databases.

## 2 Implementation

ProSeqViewer is a TypeScript library implemented to be compatible with modern applications. It is lightweight, fast, interactive, dynamic, responsive and it adapts to any screen size. It generates pure HTML which is compatible with any browser and operating system. The source code and an extensive documentation including use cases are available from the URL: <https://github.com/BioComputingUP/ProSeqViewer>.

The HTML generated by ProSeqViewer is lightweight and simple, allowing to render thousands of sequences without freezing the page. Single sequences and MSAs are treated the same way in ProSeqViewer. Each letter is encapsulated in a column element (just one cell for single sequences) which is part of a block of columns. Blocks are graphically separated to improve readability. Sequence indexing is either placed at the beginning of each line or right aligned on top of each block. By default, blocks are automatically moved to a new line when they exceed the container width. However, block wrapping can be disabled if the user wants to slide MSAs horizontally.

Sequences and MSAs can be colored, highlighted and decorated with positional annotations. Figure 1 shows an example featuring a single protein sequence and two annotated versions of the same MSA. ProSeqViewer comes with a predefined set of secondary structure and post translational modification SVG icons, which can be placed on top of the sequence. The library includes predefined color palettes, e.g. “Clustal” to highlight amino acid physico-chemical properties (Figure 1, panel B) and “Blosum” to visualize substitution scores of aligned amino acids (Figure 1, panel C). It is also possible to color regions by providing a list of start end boundaries (Figure 1, panel A) or sequence patterns by providing regular expression strings. Custom icons and color gradients can be defined by the user. Some features are specifically designed for MSAs, i.e. the calculation of a consensus sequence displayed right above the alignment. The consensus can represent either the sequence identity or the psycho-chemical conservation of MSA columns (Figure 1, panel C). Finally, ProSeqViewer implements two listeners that capture mouse interactions. The first recognizes mouse selections and stores the selected sequence (or MSA) fragments in the clipboard, e.g. for copy and paste. The second listener captures mouse double clicks and stores the corresponding position and sequence letter.

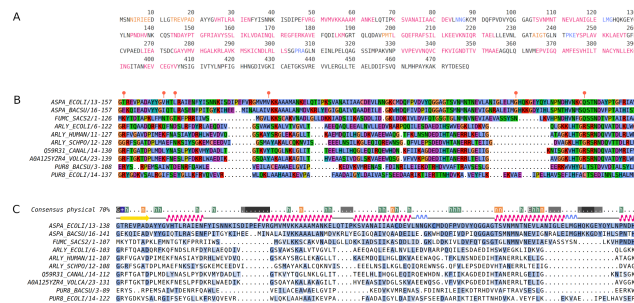


Fig. 1. Examples of ProSeqViewer output. A) Sequence of the Aspartate ammonia-lyase protein (UniProtKB P0AC38) with residue numbering displayed on top of

sequence blocks of ten residues. Residues are colored based on their secondary structure (pink helixes, blue turns and yellow strands). B) A multiple sequence alignment (MSA) of the same protein with ten homologs showing mutations as lollipop icons on top. Amino acids colored by their chemico-physical properties. C) The same MSA showing the consensus sequence (first row) and secondary structure annotations (second row). The consensus has a black background if the column is fully conserved or a symbol representing the most conserved chemico-physical property. Columns with less than 70% conservation are indicated with a dot. MSA sequences are colored with the “Blosum” color scheme to highlight the substitution score with respect to the first sequence. Sequence names and start/end positions are listed on the left.

## 3 Summary

ProSeqViewer is a tool specifically designed for both single sequences and MSAs, providing a unique solution for multiple visualizations. It offers a unique alternative to other tools by generating a lightweight HTML that renders thousands of sequences without freezing the page. Annotation icons, gradients, sizing and spacing are fully customizable. It adapts dynamically to the page width supporting visualization in small devices or minimized windows. ProSeqViewer implements listeners that capture mouse interactions and expose sequence selections to other active components of the website and with the clipboard. Input and configurations are stored in typed variables thus minimizing the chance of bugs when integrating the viewer into complex websites. The library is open-source and distributed under the CC-BY License. The code along with use-cases and extensive documentation are provided. We commit to its maintenance and welcome feature requests and feedback. ProSeqViewer is also available in the NPM registry from URL: <https://www.npmjs.com/package/proseqviewer>.

## Funding

This project received funding from the European Union’s Horizon 2020 research and innovation program [778247, 823886, 952335]; and the Italian Ministry of University and Research (PRIN 2017) [2017483NH8].

Conflict of Interest: none declared.

## References

Brown,N.P. et al. (1998) MView: a web-compatible database search or multiple alignment viewer. *Bioinformatics*, **14**, 380–381.

Gouet,P. et al. (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.*, **31**, 3320–3323.

Hatos,A. et al. (2020) DisProt: intrinsic protein disorder annotation in 2020. *Nucleic Acids Res.*, **48**, D269–D276.

Mistry,J. et al. (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.

Paladin,L. et al. (2021) RepeatsDB in 2021: improved data and extended classification for protein tandem repeat structures. *Nucleic Acids Res.*, **49**, D452–D457.

Piovesan,D. et al. (2021) MobiDB: intrinsically disordered proteins in 2021. *Nucleic Acids Res.*, **49**, D361–D367.

Tanyalcin,I. (2019) Lexicon-Mono-Seq, DOM Text Based Async MSA Viewer. *J. Open Source Softw.*, **4**, 1407.

The UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.

Waterhouse,A.M. et al. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.

Yachdav,G. et al. (2016) MSAViewer: interactive JavaScript visualization of multiple sequence alignments. *Bioinformatics*, **32**, 3501–3503.

Zahn-Zabal,M. et al. (2020) The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Res.*, **48**, D328–D334.