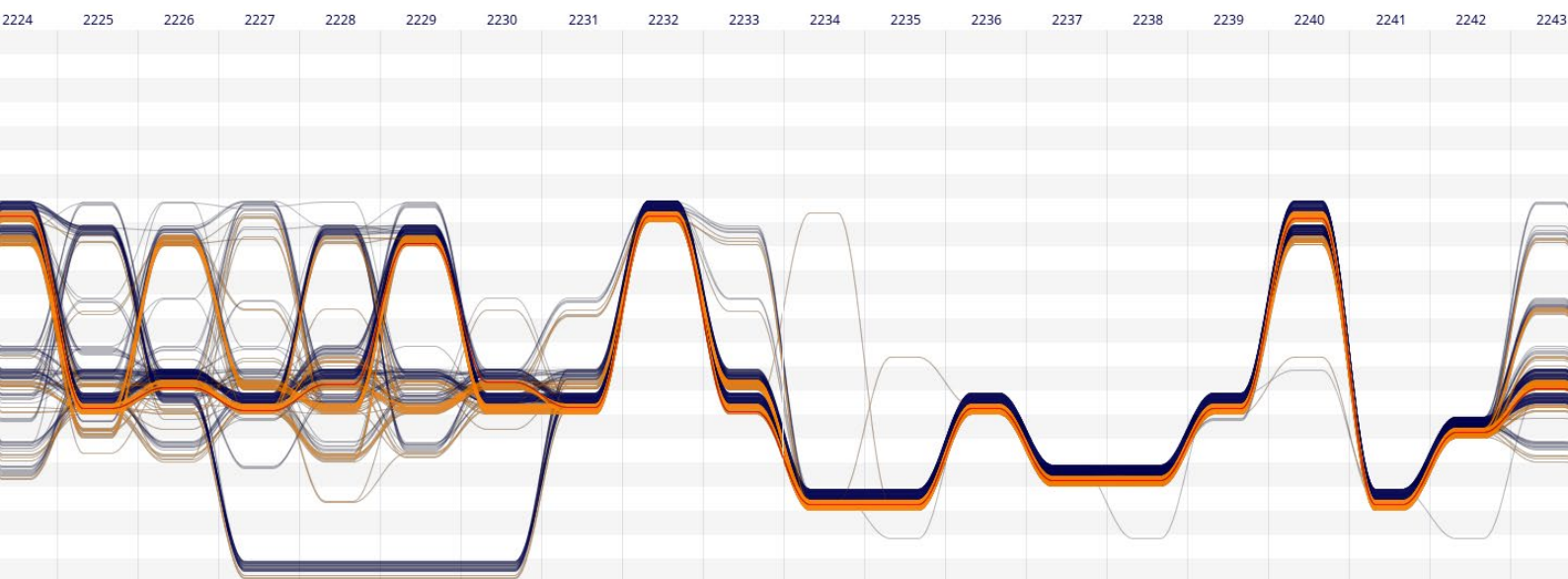


Sequence Bundles

SHORT USER MANUAL



Over the past year, Science Practice has been working with the Goldman research group at the EMBL European Bioinformatics Institute to design a new interactive tool for visualising, exploring and discovering sequence motifs using our Sequence Bundles visualisation method.

The method has been described in our recent publication in BMC Proceedings:
<http://doi.org/10.1186/1753-6561-8-S2-S8>
and in a Sequence Bundles intro video:
<http://vimeo.com/109023860>

Sequence Bundles aim to enable you to do two main things: visualise and communicate your MSA data as a set of continuous lines and explore your dataset by manipulating this visualisation. We hope that Sequence Bundles will help you generate new insight and support hypothesis making.

The Science Practice team

<http://science-practice.com>
@science-practice
af@science-practice.com

Contents

1. About the software
 - 1.1 Authors
 - 1.2 Current version
2. Package contents
3. Software requirements
 - 3.1 Windows users
 - 3.2 Mac users
4. Launching the software
 - 4.1 Windows users
 - 4.2 Mac users
5. Importing alignment
6. User workspace
7. Rendering options
 - 7.1 Line shape
 - 7.2 Gaps
 - 7.3 Colours
8. Consensus sequence
9. Y-axis properties
10. Sequence selection
11. Sequence grouping
 - 11.1 Assigning into groups
 - 11.2 Clustering into groups
12. Reference sequence
13. Exporting
 - 13.1 Export to PNG (bitmap)
 - 13.2 Export to SVG (vectors)
14. Reporting issues
 - 14.1 Feedback and recommendations
 - 14.2 Bugs and problems
15. Legal note
16. Example data files

1. About the software

Sequence Bundles is a visualisation module used in a larger software tool called Alvis.

In this manual we will provide basic instructions for using Sequence Bundles within the Alvis environment.

1.1 Authors

The sequence Bundles visualisation method was created by Science Practice, a design and research studio in London (<http://science-practice.com>).

The Alvis software was developed by Roland Schwarz at the Goldman group at the EMBL-EBI (<http://www.ebi.ac.uk/research/goldman/members>).

1.2 Current version

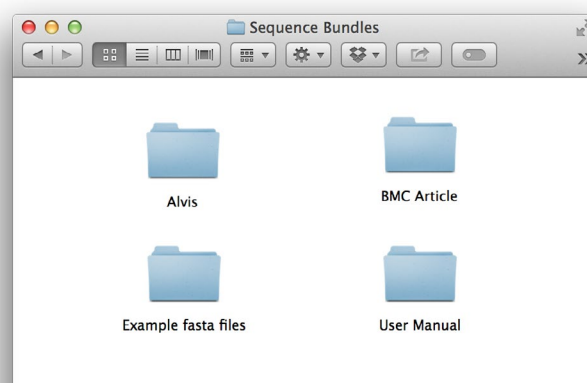
Sequence Bundles and Alvis have been so far developed as an alpha version for user studies with bioinformatics experts. The software is still imperfect

and may crash or not respond to user input as anticipated. Please keep this in mind when using Sequence Bundles. Thank you for your help.

2. Package contents

Inside the Sequence Bundles package you will find the following elements:

1. Alvis software tool,
2. Example fasta files,
3. Sequence Bundles intro video,
4. This user manual.



3. Software requirements

3.1 Windows users

- Java Runtime Environment 7 Update 67,
- ~200 MB free hard drive space,
- minimum 2 GB RAM (4GB recommended),
- Windows XP or higher,
- Minimum screen width 1080p.

3.2 Mac users

- Java Runtime Environment 7 Update 67,
- ~200 MB free hard drive space,
- minimum 2 GB RAM (4 GB recommended),
- OS X version 10.9.4 or older (newer OS X versions not fully supported),
- Not supported for computers with retina displays,
- Early 2011 model recommended, if used on MacBook Pro,
- Minimum screen size 13".



4. Launching the software

4.1 Windows users

In order to launch Alvis you may need to provide all necessary permissions for your computer to open a downloaded application.

4.2 Mac users

When launching Alvis for the first time on a Mac, you may be required to provide a one-off permission for your system to launch an application developed by a party other than Apple or identified developers. Whether you will receive the following request depends on your system security settings.

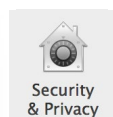


In case you are presented with this warning, please go to your



System Preferences menu

and then to the



Security & Privacy section

where you will be able to allow for Alvis to be launched on your Mac.

5. Importing alignment

To start exploring your data with Alvis, import a selected alignment using the “Import alignment...” command.

[File > Import alignment...](#)

You will be able to select your own data files (alignment in fasta format) or one of the examples provided (see “Example fasta files” package folder or inside the “Alvis/dist/examples” directory

6. User workspace

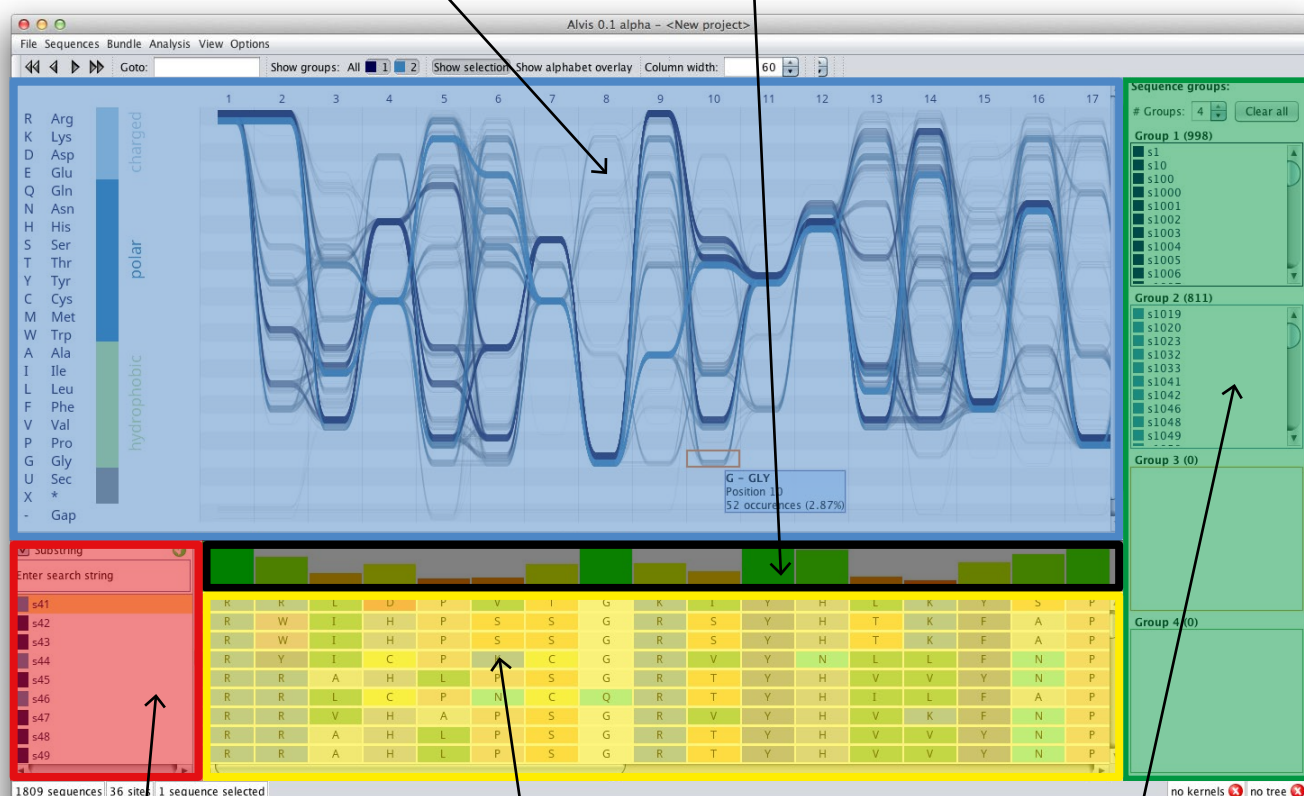
In order to overcome the shortcomings of a single way of representing your data, Alvis allows you to view your data in a number of ways simultaneously. Alvis uses a visualisation strategy called “linked views”, which allows you to select points of interest in one view and

see them highlighted across all viewing modes in the software.

The image below outlines all elements of Alvis user workspace, which we will refer to later in this manual.

Sequence Bundles view

Conservation/Sequence Logos view



MSA list

Alignment view

Groups view

7. Rendering options

The Sequence Bundles visualisation method provides a range of rendering options which influence the appearance of visualised sequences. This can help you explore your data in a variety of ways, enable the noticing of unexpected features and allow to direct your visual investigation of the dataset.

To adjust rendering settings, change colours assigned to groups and define the preferred method for representing alignment gaps, go to:

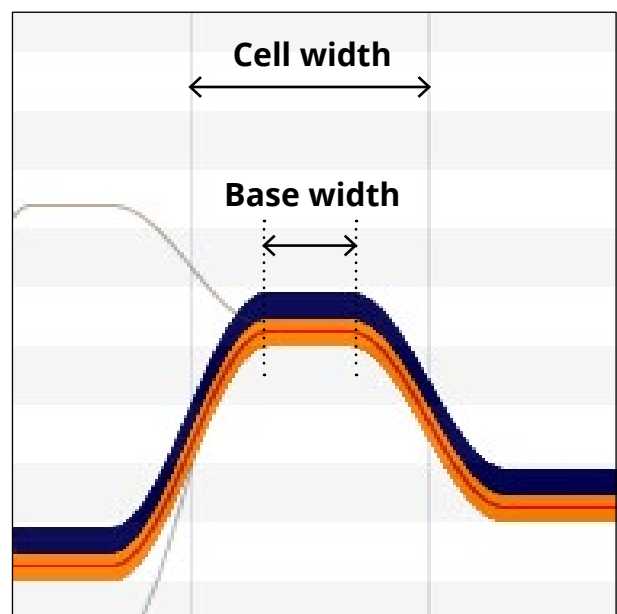
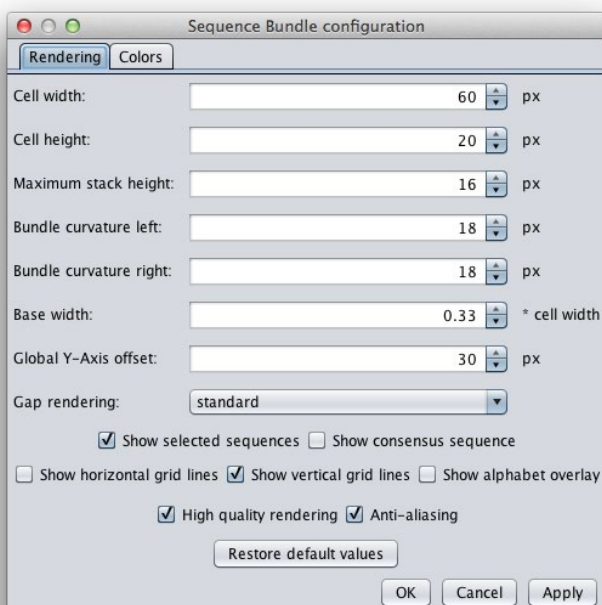
[Options > Rendering options...](#)

7.1 Line shape

In Sequence Bundles the shape of lines and the structure of the bundles can be controlled with a range of variables. The most important are: “Cell width” which defines the width of each cell in the matrix in which the bundles are drawn; and “Base width” which defines

the fraction (in %) of each cell’s width that is occupied by the flat horizontal base of the line.

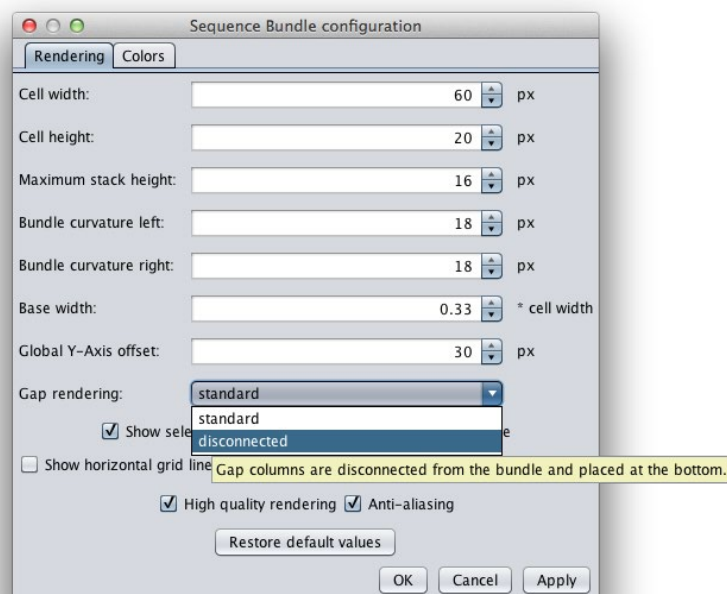
Experiment with these and with other variables to see what they control. Use “Apply” button to view instant changes.



7.2 Gaps

Sequence Bundles provide two ways of encoding gaps used for aligning sequences in MSAs. The standard way of representing gaps treats them as any other residues in the bundle; therefore, in the alignment, there are vertical curves linking amino acids to the gaps.

The second method for representing gaps treats them in disconnection from other residues, avoiding the display of vertical linking curves (which in some cases may clutter the view). Gaps are deposited on their own at the bottom of the diagram. You can choose the best method for visualising your data.



7.3 Colour

Sequence Bundles allow the display of sequences allocated to distinct groups. This enables easy and quick contrasting and comparison of, for example, different organism families or types of sequences.

You can assign selected colours to groups of sequences or other rendering elements by choosing the "Colors" tab after selecting the "Rendering options..." command from the "Options" menu.



8. Consensus sequence

Sequence Bundles can be explored in reference to the consensus sequence.

[Bundle > Show consensus](#)

To display the consensus sequence, select the “Show consensus” command.

9. Y-axis properties

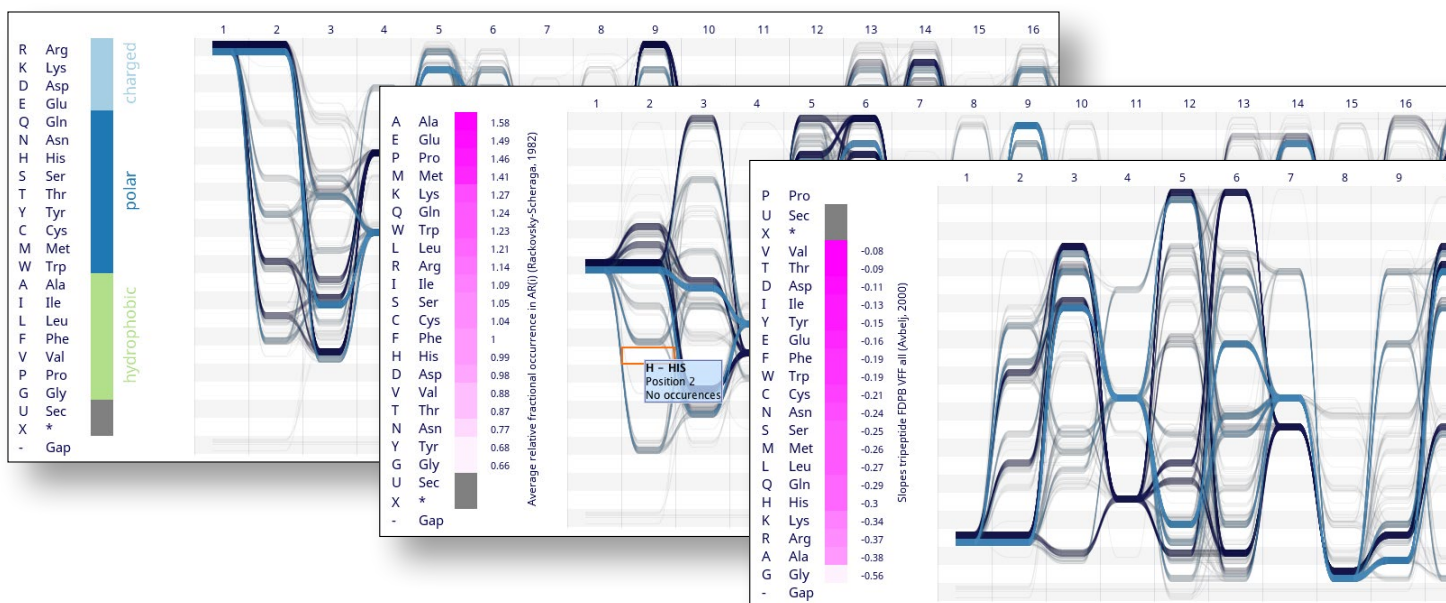
The vertical Y-axis in Sequence Bundles is highly meaningful. By default, the residues listed on the Y-axis are divided into three main categories: charged, polar and hydrophobic.

You can reorganise the Y-axis according to given properties selected from the AAIndex (e.g. hydrophobicity, bulkiness, isoelectric point, normalized flexibility

parameters, refractivity, residue volume or side chain volume amongst many more).

To select a preferred organisation for the Y-axis choose the Legend command.

[Bundle > Legend > From AAIndex](#)



10. Sequence selection

Alvis offers a number of methods for selecting sequences in Sequence Bundles. The simplest is to click on a given line to select it. The selection highlights.

Press SHIFT to make an additive selection using the cursor.

You can also use a range of commands available in the “Sequences” menu to refine your selection.

[Sequences > Select all](#)

[Sequences > Select by group > ...](#)

[Sequences > Deselect all](#)

[Sequences > Invert selection](#)

[Sequences > Hide selected](#)

[Sequences > Show selected](#)

[Sequences > Show all](#)

11. Sequence grouping

Sequence Bundles allow the display of sequences allocated to distinct groups. This enables easy and quick contrasting and comparison of, for example, different organism families or types of sequences.

There are two main ways in which sequences can be assigned into groups:

1. manual assigning into groups,
2. automatic clustering into groups.

11.1 Assigning into groups

Sequence Bundles allow you to assign different sequences into groups. In order to do this, select the sequences you want to move into a separate group and option-click on one of the highlighted sequences in the MSA list view. In the pop-up menu select “Assign group...” and select the group into which the selection should be assigned to.

You can observe and manage sequence groups in the Group view. By default, this is not visible in Alvis. In order to open it, select the “Groups” command from the “View” menu.

[View > Groups](#)

11.2 Clustering into groups

Alvis can also automatically allocate imported sequences to a specified number of groups that will best match their similarities and differences. In order to cluster sequences into groups automatically, first compute the alignment kernel and then select the “Cluster sequences into groups...” command.

First...

[Analysis > Compute alignment kernel...](#)

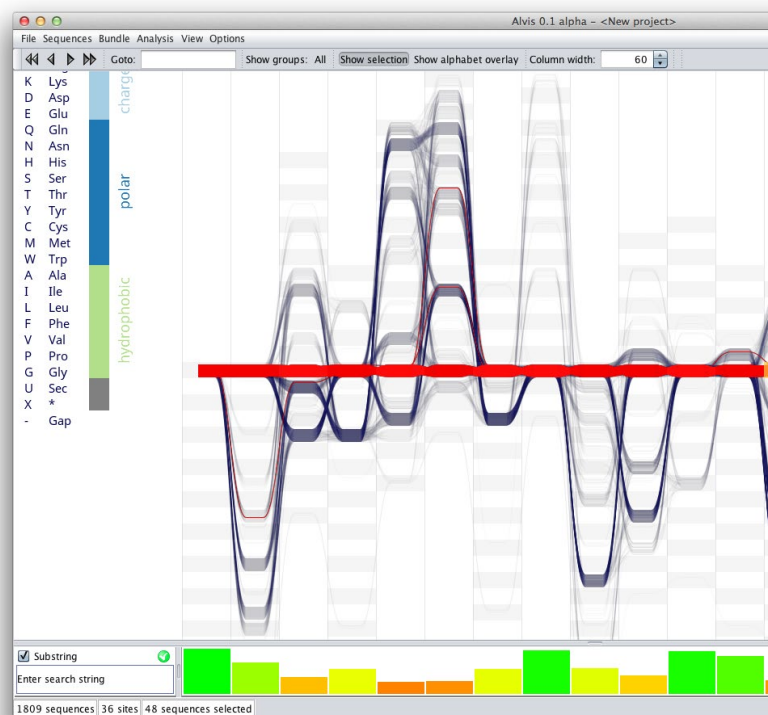
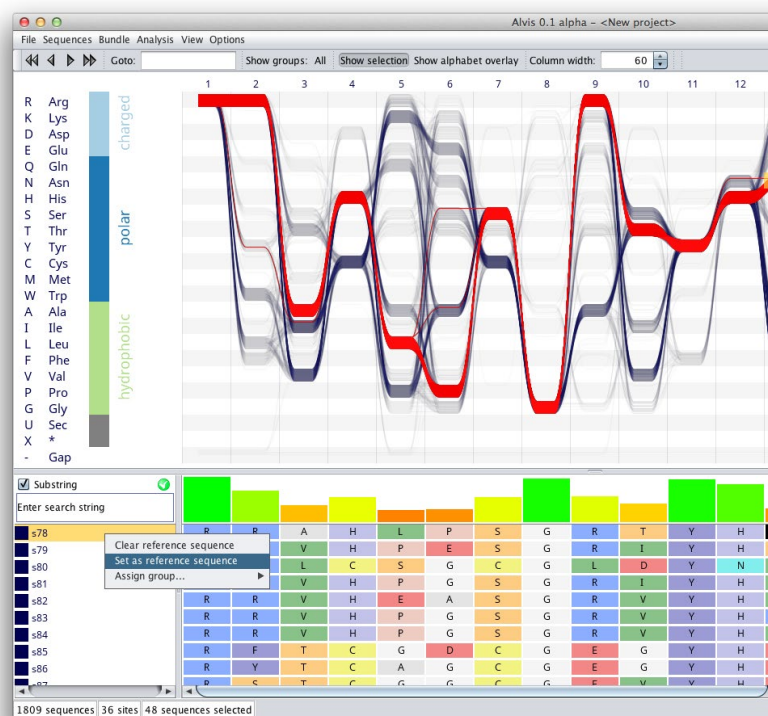
...and then

[Analysis > Cluster sequences into groups...](#)

12. Reference sequence

Sequence Bundles can be organised according to physiochemical properties (see 9. Y-axis properties) but they can also be organised in respect to a selected sequence that can act as reference for other sequences in the alignment. This enables advanced comparison of sequences.

In order to assign a reference sequence, select the single sequence or a number of sequences you want to use as reference and option-click on one of them highlighted in the MSA list view. In the pop-up menu select “Set as reference sequence”.



13. Exporting

13.1 Exporting to PNG (bitmap)

It is possible to export Sequence Bundles as a bitmap image that can later be used to present findings to other researchers or in publications.

In order to export a PNG image of the rendering, choose the “Export to file...” command, then select the PNG format and your preferred resolution in the export window menu.

[Bundle > Export to file...](#)

13.2 Exporting to SVG (vector)

It is also possible to export Sequence Bundles as vector artwork that can be later explored or optimised visually in a vector graphics programme.

In order to export an SVG artwork file choose the “Export to file...” command, then select the SVG format and set the resolution settings to **72dpi** in the export window menu.

[Bundle > Export to file...](#)

14. Reporting issues

14.1 Feedback and recommendations

We are really interested in learning more about how the Sequence Bundles visualisation and tool can help support researches in their work. For this purpose, we would be very grateful if after trying out the tool you would be available for a short telephone or Skype call to share with us your thoughts.

We are primarily interested in the following two aspects of the tool:

- Discovery: Does Sequence Bundles enable the discovery of interesting features and motifs? Does it help generate new insights into your data and research?
- Usability: How can we improve your experience of using the tool? Are there any additional features or general observations you would like to suggest?

We also welcome any additional feedback and recommendations at: af@science-practice.com

14.2 Bugs and problems

As we're still at an early stage in the development of the tool, we are aware that there might be some bugs in the code. We would really appreciate it if you could take the time to let us know if anything goes wrong with the software so that we can get it running smoothly.

If you encounter an error, could you please send us an email at mk@science-practice.com with the following details:

- Short summary of the bug,
- Steps to reproduce the bug,
- Result of performing the steps,
- Expected result of performing the steps,
- Hardware platform used,
- Operating system used,
- Snapshot of the bug,
- Plus any attachments that can help us track down the bug and fix it.

15. Legal note

We would like to remind you that Sequence Bundles, as part of the Alvis visualisation software, is in an early development stage. Therefore, the software may be unstable and could cause crashes or data loss. Science Practice and the Goldman Group are not able to take any responsibility in the case of these events and it is advisable that you close any other applications when using the software.

Also, as we are still working on the tool, we would really appreciate, if you didn't share the software directly with anyone else. If you know of someone who might be interested in trying out Sequence Bundles, please let us know and we will be happy to get in touch and share the tool with them.

If you are interested in using Sequence Bundles for publishing purposes, please let us know so that we can provide you with additional advice on how to obtain a high quality image and adequately reference the method.

16. Example data files

The “Example fasta files” folder included in the Sequence Bundles package that you have downloaded contains a number of MSA files of various origin.

The “ADKlid_1809sequences” file is the dataset kindly provided by Drs. Magliery and Sullivan at The Ohio State University for the purposes of the BioVis 2013 Contest. The MSA contains 1809 sequences of the bacterial protein ADK Lid domain.

The “mammal_mtCDNA”, “oshea_ii”, “phosphofructokinase_muscle”, “PB2”, “kinase_cyclase_peptides” as well as “All_long_selected4.mus” fasta files are kindly provided by the EMBL-EBI.

There is also an additional file titled “211_Curated_sequences_combined” which is a fictional alignment prepared by Science Practice for the purposes of Alvis testing only. It is based on parts of insulin protein.

Thank you again for using Sequence Bundles. We hope you enjoy the visualisation tool and find it helpful.